

SAG' MIR, WAS ICH NOCH NICHT WEISS

—

INDIVIDUALISIERUNG DES WISSENSERWERBS ÜBER ADAPTIVE AUSWAHL DER LERNINHALTE MITHILFE DER NONMETRISCHEN MULTIDIMENSIONALEN SKALIERUNG UND PROKRUSTES-TRANSFORMATION

Abhandlung
zur Erlangung der Doktorwürde
der Philosophischen Fakultät
der Universität Zürich

vorgelegt von
Roland Streule
Appenzell (AI)

Angenommen auf Antrag von
Prof. Dr. Damian Läge und
Prof. Dr. Wolfgang Marx

Zentralstelle der Studentenschaft
Zürich 2007

Inhaltsverzeichnis

Zusammenfassung.....	1
Abstract.....	1
1. Einleitung	3
1.1. Theoretischer Überblick und Problemstellung.....	5
1.2. Stand der Dinge.....	8
1.3. Aufbau der Arbeit und Einordnung in das Gesamtkonzept.	9
1.4. Gesamtdiskussion der Resultate.....	11
1.5. Literatur.....	14
2. Die Veränderung von Wissenskarten beim Merkmalslernen	17
Zusammenfassung.....	19
2.1. Einleitung	19
2.2. Experiment 1	23
2.2.1. Fragestellung und Hypothesen.....	23
2.2.2. Methode und Durchführung.....	23
2.2.3. Resultate.....	26
2.2.4. Diskussion.....	33
2.3. Experiment 2	34
2.3.1. Fragestellung und Hypothesen.....	34
2.3.2. Methode und Durchführung.....	34
2.3.3. Resultate.....	36
2.3.4. Diskussion.....	40
2.4. Gesamtdiskussion.....	42
2.5. Literatur.....	43
3. Der NMDS-Stresswert als Maß für Assimilation und Akkomodation beim Lernen von Wissensstrukturen	45
Zusammenfassung.....	47
3.1. Einleitung	47
3.2. Methode.....	56
3.2.1. Experiment 1	56
3.2.2. Experiment 2	59
3.2.3. Auswertung der Daten und erwartete Befunde	60
3.3. Resultate	61
3.4. Diskussion	66
3.5. Literatur.....	69

4. Der ObjLoss einer Prokrustes-Transformation als Maß für Wissen.....	73
Zusammenfassung.....	75
4.1. Einleitung	75
4.2. Simulation	80
4.2.1. Methode	80
4.2.2. Resultate.....	83
4.2.3. Diskussion.....	86
4.3. Experiment	87
4.3.1. Methode	87
4.3.2. Resultate.....	92
4.3.3. Diskussion.....	96
4.4. Gesamtdiskussion.....	98
4.5. Literatur	101
5. Effizienzsteigerung in der ähnlichkeitsbasierten Wissensdiagnostik.....	103
Zusammenfassung.....	105
5.1. Einleitung	105
5.2. Methode.....	112
5.3. Resultate	115
5.4. Diskussion	121
5.5. Literatur	125
6. Adaptive Lernempfehlungen über Prokrustes-Transformation und Kohonennetz	127
Zusammenfassung.....	129
6.1. Einleitung	129
6.2. Methode.....	137
6.3. Resultate	140
6.4. Diskussion	142
6.5. Literatur	149
7. A Structure-Based Expert Model of the ICD-10 Mental Disorders	151
Abstract	153
7.1. Introduction	153
7.2. Methods	154
7.3. Results	156
7.4. Discussion	161
7.5. References	164

8. Das Zürcher Wissensstrukturmodell für Psychopathologie (ZWP) – eine Pilotstudie	167
Zusammenfassung.....	169
8.1. Einleitung	169
8.2. Methode.....	173
8.3. Resultate	175
8.4. Diskussion	177
8.5. Literatur	179
9. Pilot-INTUS	181
Zusammenfassung.....	183
9.1. Einleitung	183
9.2. Methode.....	189
9.3. Resultate	193
9.3.1. Überprüfung der Voraussetzungen	193
9.3.2. Adaptives Lernen vs. Standard-Lernen.....	196
9.3.3. Prognosequalität der ObjLoss-Werte	199
9.4. Diskussion	201
9.5. Literatur	206
Danksagung.....	209
Curriculum Vitae	211

Zusammenfassung

Computergestützte Lehr- und Lernprogramme bieten einen innovativen Zugang, den sich neu präsentierenden bildungspolitischen Anforderungen gerecht zu werden. Individuelle, auf den Lerner zugeschnittene Wissensvermittlung ermöglicht eine effiziente und flexible Wissensakkumulation. Intelligente Tutorielle Systeme (ITS) verfügen über Lehrstoffexpertise, womit individualisiert auf den aktuellen Wissensstand einer Person Rücksicht genommen werden kann. Die Gewährleistung der inhaltlichen Adaptivität solch eines Systems erfordert eine valide Wissensdiagnostik. Das Methodeninventar sieht sich dabei mit zwei kompetitiven Schwierigkeiten konfrontiert: Zum einen soll der Wissenstest umfassend und exakt, zum anderen möglichst effizient sein. Die vorliegende Arbeit stellt eine Verfahrenslogik vor, welche mithilfe von Kognitiven Karten strukturelles Faktenwissen zu messen in der Lage ist. Die Nonmetrische Multidimensionale Skalierung (NMDS) und die Prokrustes-Transformation bilden die Eckpfeiler der Diagnoseprozedur, mithilfe dessen nicht nur eine Vorhersage bezüglich des Sachwissens, sondern darüber hinaus relationale Gefüge, Integration und Elaboration der Gedächtnisinhalte visualisiert und einer qualitativen Beurteilung zugänglich gemacht werden können. Experimentelle und simulative Studien untersuchen die Vorhersageperformance der Verfahrenslogik und geben darauf aufbauend Empfehlungen für den praktischen Einsatz in automatisierten Lehr- und Lernsystemen. Die Resultate führen zudem zu einem didaktischen Konzept, welches übereinstimmend mit Modellen der Pädagogischen Psychologie einen assimilativ orientierten Lernprozess propagiert.

Abstract

Computer-aided teach- and learning-programs provide an innovative access to satisfy modern educational requirements. Individual and individually adapted mediation of knowledge offers an efficient and flexible accumulation of knowledge. Intelligent Tutorial Systems (ITS) possess teaching material expertise, that can take into consideration the current state of knowledge of a person in an individualized way. The guarantee of adaptivity in respect of content of such a system requires valid knowledge diagnostics. Methodological inventory is confronted with conflicting issues in this case: On the one hand knowledge testing should be all-inclusive and precise, on the other hand as efficient as possible. The present study presents a methodological way to measure structural factual knowledge by means of Cognitive Maps. Nonmetric Multidimensional Scaling (NMDS) and Procrustean Transformation are the principal components of the diagnostic procedure. These methods allow a visualization of a prediction of factual knowledge as well as of the relational structure, integration and elaboration of memory content and offer access to a qualitative evaluation. Experimental and simulation studies investigate the prediction performance of methodology and offer according recommendations for practical application in automated teach- and learning systems. Furthermore, the results lead to a didactical concept, which propagates – in agreement with models of pedagogical (educational) psychology – an assimilative oriented learning process.

1.

Einleitung

1.1. Theoretischer Überblick und Problemstellung

Wissen wird in der heutigen Informationsgesellschaft vermehrt als *das* zentrale Gut definiert. Das „Know How“ einer Gesellschaft kommt einer Visitenkarte gleich, welches nicht erst seit Einführung der Pisa-Studien immer weiter verbessert werden muss. Es entwickeln sich immer neue und komplexere Informationssysteme, welche einen Austausch und Zuwachs von Wissen in immer kürzerer und umfassender Form verlangen. Das Schlagwort „lebenslanges Lernen“ kann als das zentrale Paradigma betrachtet werden, welches aufgrund dieser ständig wachsenden Vernetztheit seine Relevanz im Sinne der vielerorts geforderten Interdisziplinarität in der Wissensgenerierung erfährt. Diesem Umstand müssen auch Bildungsinstitutionen gerecht werden, indem sie ihre Bildungsprogramme den neuen Gegebenheiten anpassen. Der klassische Frontalunterricht, wie ihn wohl noch viele Zeitgenossen am eigenen Leib erfahren durften, weicht zusehends einer immer stärkeren Individualisierung des Lernvorgangs. Den früher noch weit verbreiteten Werdegang eines einmal gelernten Berufs, den man dann zeitlebens ausübte, findet man mittlerweile wohl kaum mehr in individuellen Biographien. Aufmerksam das Wechselspiel zwischen Angebot und Nachfrage im schnelllebigen Arbeitsmarkt beobachtend, muss das Individuum Flexibilität beweisen, um in der beruflichen Lebensplanung nicht plötzlich vor verschlossenen Türen zu stehen. Den Aus-, Weiter- und Fortbildungsinstitutionen bleibt somit die schwierige Aufgabe zu lösen, Wissen so effizient und qualitativ hochwertig zu vermitteln, dass ein Optimum zwischen den Polen der Spezialisierung und des Weitblicks innerhalb der einzelnen Lernprozesse gefunden werden kann.

Die angesprochene Forderung nach Individualisierung von Lernvorgängen wird schon heute verstärkt über neue Formen der Kommunikation angegangen. Die Umsetzung von Wissensvermittlung über computergestützte Ausbildungseinheiten erfährt seit mehreren Jahren einen regelrechten „Boom“ und ergänzt klassisch strukturierte Bildungsinstitutionen in (mindestens) zweifacher Hinsicht: Die virtuelle Vernetztheit ermöglicht es, das Lernen orts- und zeitunabhängig zu gestalten, indem beispielsweise über das Internet Lernprogramme zu jeder Tageszeit und von irgendeinem Standpunkt aus absolviert werden können. Des Weiteren ermöglicht die Computerisierung die Entwicklung intelligenter Software für einen dynamischen und optimal auf den Lerner zugeschnittenen Wissensvermittlungsapparat.

Intelligente Tutorielle Systeme¹ (ITS) bilden einen Ansatz zur Gewährleistung der an den Wissensstand des Lerners angepassten schulischen Bildung, wobei die Idee von Lehrmaschinen nicht neu ist. Leutner (1992a) stellt in seinem Buch „Adaptive Lehrsysteme“ einen historischen Abriss über die Entwicklung solcher Unterrichtsformen dar, welche, angefangen mit ersten Ansätzen in den zwanziger Jahren des letzten Jahrhunderts mit Maschinen zur „Intelli-

¹ Auf definitorische Unterscheidungen zwischen den unzähligen und zum Teil gleichbedeutenden Bezeichnungen für (virtuelle) Lehrsysteme wie bspw. ITS (Intelligente Tutorielle Systeme) und ATS (Adaptive Tutorielle Systeme) wird hier nicht weiter eingegangen. In der vorliegenden Arbeit wird der etwas allgemeinere und geläufigere Terminus ITS verwendet und meint einfach ausgedrückt, dass sich das Lehrsystem an die individuellen Bedürfnisse des Lernenden anpassen kann. Im Speziellen bedeutet dies für diese Arbeit, dass das System auf individuelle Wissensdefizite entsprechend didaktisch und inhaltlich reagiert.

genzprüfung und Informationsvermittlung“ (Pressey, 1926) und deren Weiterentwicklung durch Skinner (1958) und Crowder (1959) in den sechziger Jahren in den ersten computerisierten Systemen mündeten (CAI: computer assisted instruction). Erst mit dem Aufkommen der Künstlichen Intelligenz aber wurden an die CAIs neue Anforderungen gestellt, nämlich, dass sich die nun als Intelligente Tutorielle Systeme bezeichneten Programme mehr und mehr dem Instruktionsverhalten einer menschlichen Lehrkraft angleichen (s. Goldstein, 1982), also über Lehrstoffexpertise verfügen. Erst der Vergleich zwischen dem Ist-Zustand des Lerners und dem Soll-Zustand des Expertenmodells führen zu „intelligentem Verhalten“ des Systems und bilden die Basis für einen ökonomischen und individuell angepassten Lernvorgang (vgl. zu den Komponenten eines ITS Van Merriënboer, Van Den Berg & Maaswinkel, 1989).

Der gewöhnliche Klassenunterricht nimmt lediglich auf den Kenntnisstand aller Lernenden Rücksicht – ein Umstand, der sich bei den heute zu beobachtenden „Großklassen“ in Massenanveranstaltungen an Universitäten oder auch auf Grundstufenniveau noch verschärft – und verlangt von der Lehrkraft ein hohes Maß an Kontrolle, Flexibilität und didaktischer Kompetenz. Virtuelle Unterrichtsformen demgegenüber erlauben eine stärkere Hinwendung auf das aktuelle Wissen des Einzelnen und ermöglichen die Chance, über eine effiziente Wissensmessung automatisiert eine individuell angepasste Lerneinheit zusammenzustellen. Ein derartiger „adaptiver Unterricht“ muss zwei Zielen genügen: Zum einen der Fokussierung auf die Gestaltung didaktisch optimal aufbereiteter multimedialer Unterrichtsinhalte. Zum anderen der überaus wichtigen, verlässlichen und effizienten Statusdiagnose als Grundlage für die Erstellung der kommenden Lektion. Eine lernprozessbegleitende Diagnose des Wissens führt jedoch zu einem Dilemma, auf das bereits Leutner (1992b, 1993) hingewiesen hat: Auf der einen Seite soll eine möglichst umfassende und genaue Diagnose erstellt werden, um bei der Zusammenstellung der folgenden Lektionen so adaptiv wie möglich sein zu können. Auf der anderen Seite will man aber nicht zuviel Zeit und kognitive Ressourcen der Lernenden für die Diagnose beanspruchen, weil beides von der verbleibenden Lernzeit abgeht. Die Adaptivität im Inhalt gewinnt ihre Relevanz dadurch, dass zum einen Mehrfachlernen weitgehend vermieden und zum anderen, dass falsches Wissen (welches über einen selbstgesteuerten Lernprozess nur schwerlich detektiert wird, vgl. Charney, Reder & Kusbit, 1990 und Strohschneider, 1990) ausgemerzt werden kann.

Klein (2000) resümiert in diesem Zusammenhang die Schwierigkeiten einer solchen Adaptivität. Er führt aus, dass die *Adaptierbarkeit* (d.h. die Anpassbarkeit des Systems durch den Nutzer, wie z.B. die Wahl unterschiedlicher Bildschirmarrangements oder das An- und Ausschalten von Hilfefunktionen) technisch wie konzeptionell wesentlich leichter zu lösen ist als die inhaltliche *Adaptivität* des Lernmaterials (vgl. auch Leutner, 1992a, 1995). In der Tat beschränken sich die heutigen ITS-Systeme weitgehend auf die Vermittlung regelbasierten Wissens, Wissen also, welches sich mit Wenn-dann-Regeln charakterisieren lässt (als Umsetzungs-Beispiele für Mathematik- und Physikwissen s. Lukas & Albert, 1999 und Siemer & Angelides, 1998; für einen erweiterten Überblick über bestehende ITS-Systeme s. Daub, 2001) und vielfach durch einen kleinen Satz an Entscheidungs- und Transformationsregeln abgedeckt werden kann. Das Effizienzproblem in der Wissensüberprüfung wird jedoch besonders deutlich bei der Diagnostik merkmalsbezogenen Sachwissens. Geht es um die unzäh-

ligen, für die Qualität von Expertenentscheidungen aber zwingend notwendigen Fakten, müsste für eine zuverlässige Diagnose der gesamte Wissensstand innerhalb eines Fachgebietes mit Hunderten von Fragen abgetestet werden. Nur so wäre man in der Lage, Lücken und Falschwissen einer Person zu detektieren, was in der Praxis ein nicht durchführbares Unterfangen darstellt.

Ein zweites zentrales Problem besteht darin, dass die Elemente des Wissens nicht einfach im Gedächtnis unverbunden nebeneinander stehen. Vielmehr bilden sie – gleich auf welchem Wissensstand – eine Art System, das heißt die Elemente konstituieren unwillkürlich einen Satz von Relationen. Messmethoden für den Bereich des Sachwissens müssen demzufolge dem Umstand Rechnung tragen, dass sich Gedächtnisinhalte durch Differenziertheit, Vollständigkeit und Komplexität auszeichnen, als weit über eine einfache Anhäufung unverbundener Fakten hinausgehen (vgl. Mandl & Fischer, 2000). Eine stichprobenartige Ziehung von Wissensfragmenten, wie das in herkömmlichen Prüfungsverfahren in vielen Bildungsinstitutionen für das Bestehen eines Qualifikationskriteriums umgesetzt wird, entschärft die beschriebene Problematik nur bedingt, wenn sie das „Testlängendilemma“ (Leutner, 1992b) lediglich nach der einen Seite hin auflöst, nämlich die der Effizienz.

In den nun folgenden Kapiteln werden Arbeiten vorgestellt, die den angesprochenen Diagnostik-Problempunkten auf den Grund gehen und Lösungsansätze anbieten. Es sollen Erkenntnisse über eine Verfahrenslogik geschaffen werden, welche für den Einsatz in (computerbasierten) Lehr- und Lernsystemen die diagnostische Grundlage liefern. In letzter Konsequenz steht der Wunsch, den lästigen und mühsamen, aber unumgehbaren Werdegang von Laien zum Experten zu vereinfachen und effizienter zu gestalten.

Lösungsansatz. Die Forschergruppe der Angewandten Kognitionspsychologie der Universität Zürich hat sich zum Ziel gesetzt, die Grundlagen zur Entwicklung eines Intelligenten Tutoriellen Systems im Bereich merkmalsbasierten Sachwissens zu entwickeln. Der Fokus liegt dabei auf der *umfassenden und effizienten Wissensdiagnostik* (Statusdiagnose und Qualitätsbeurteilung) als notwendige Voraussetzung für eine adaptive Wissensvermittlung und weniger auf instruktionspsychologischen und pädagogischen Forschungsansätzen. Die Beantwortung der spezifischen Forschungsfragen sollen es ermöglichen, ein computergestütztes Intelligentes Tutorielles System für Faktenwissen mit dem Namen INTUS zu schaffen (Abbildung 1), welches in der Folge kurz umschrieben wird.

Relationale Beurteilungen von Ähnlichkeiten zwischen einem Satz an Objekten dienen als Grundlage zur Erstellung eines Wissensmodell des Lernalers (❶). Diese Wissensstrukturen des Lernalers werden mit einem systeminternen Normmodell (Expertenmodell) auf Übereinstimmung verglichen (❷). Die gefundenen Abweichungen zwischen IST und SOLL geben Hinweise auf Wissensdefizite im Kenntnisstand des Lernalers. Das Feststellen von Wissensmängeln wird dazu verwendet, die entsprechenden Informationssätze aus einer Wissensdatenbank auszuwählen (❸), für die kommende Lektion didaktisch aufbereitet zusammenzustellen (❹) und dem Lernaler zu präsentieren (❺). Nach Abschluss der Lerneinheit erfolgt eine erneute Wissensdiagnose (❻) und der sequentiell gestufte Lern- und Diagnoseprozess beginnt von

neuem (Für eine ausführliche Erläuterung der einzelnen Programmmodule vgl. Läge, 2001 und Daub, 2001).

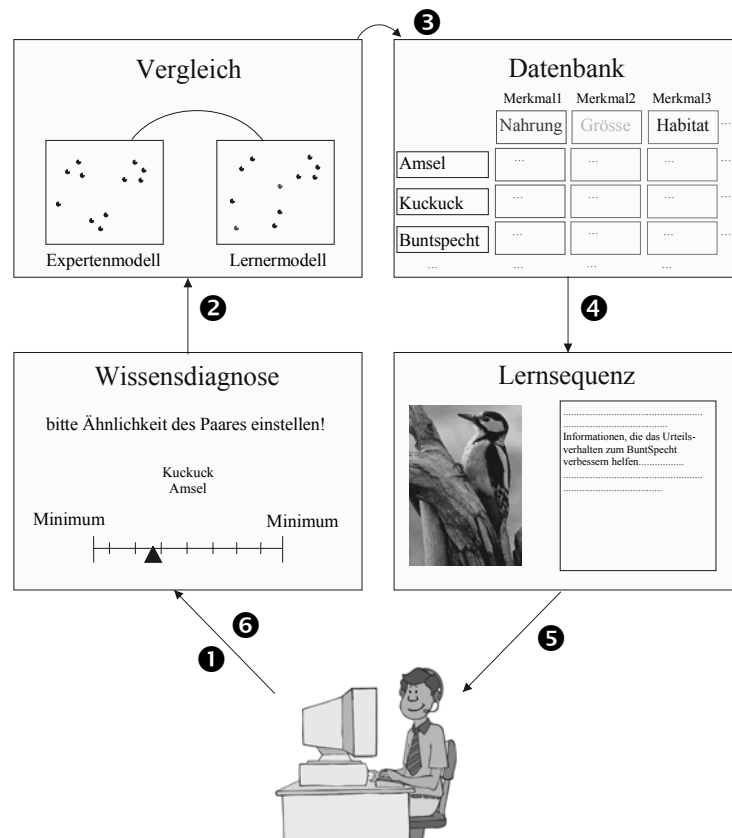


Abbildung 1. Schematischer Aufbau von INTUS (aus Daub, 2001).

1.2. Stand der Dinge

Da in den einzelnen Ähnlichkeitseinschätzungen die strukturelle Ordnung nur implizit vorhanden ist, geschieht die Umsetzung der relationalen Globalurteile in die für die Abschätzung des Lernerwissens benötigten Wissensmodelle über die *Nonmetrische Multidimensionale Skalierung* (kurz: NMDS. Die NMDS ist ein Verfahren zur Strukturanalyse und dimensionalen Reduktion multivariater Datensätze, vgl. Borg & Groenen, 1997), welche der aus der Wissenspsychologie stammenden Idee einer graphischen Abbildung von kognitiven Repräsentationsformen folgt (vgl. Mandl & Fischer, 2000). Herkömmliche NMDS-Verfahren wie beispielsweise MINISSA (Roskam & Lingoos, 1970) oder ALSCAL (ausführlich beschrieben bei Young & Lewyckyj, 1979) genügten jedoch den Anforderungen an eine auf personenspezifischer Ebene darzustellende Wissensrepräsentation nicht. Es mussten zuerst neue Auswertalgorithmen entwickelt werden, die ausreichend robust gegen Ausreißer in den Daten sind, um individuell interpretierbare Wissenskarten zu erstellen. Der neu entwickelte Algorithmus trägt den Namen RobuScal (Läge, Daub, Bosia, Jäger & Ryf, 2005).

Die Abschätzung der Qualität eines Lernermodells erfolgt über den Vergleich der Lernerkarte mit einem Expertenmodell. Die so genannte *Prokrustes-Transformation* (vgl. Gower & Dijkstra, 2004) erlaubt das Erkennen von Gemeinsamkeiten und Unterschieden, indem

die Wissens-Karten auf strukturelle Übereinstimmung überprüft werden. Auch in diesem Zusammenhang mangelte es den bereits bestehenden Verfahren an Robustheit (eine Interpretation war nur auf gemittelter Ebene möglich) und demzufolge musste eine modifizierte Prozedur erstellt werden (Läge, 2001). Auf einzelne bereits abgeschlossene Forschungsarbeiten und deren Beantwortung wie bspw. die für eine stabile NMDS-Lösung benötigte Datenmenge oder lern- und gedächtnistheoretischen Herleitungen in Bezug auf Globalurteile wird hier nicht weiter eingegangen. Für eine ausführliche und weiterführende Beschreibung der mathematischen Konzepte, der spezifischen technischen und psychologischen Detailfragestellungen sei der Leser auf Läge (2001) und Daub (2001) verwiesen.

1.3. Aufbau der Arbeit und Einordnung in das Gesamtkonzept².

Aufbauend auf den Arbeiten von Läge (2001) und Daub (2001) versucht die vorliegende Arbeit viererlei: Als erstes die eingehende Bestätigung der Validität von NMDS-Karten als Instrument zur Wissensmessung (A). Zum zweiten das Ausloten der Leistungen und Grenzen der Verfahrenslogik im Zusammenhang mit dem Entdecken von Wissensdefiziten (B). Zum dritten die Erweiterung und Spezifizierung des Diagnostikkonzepts (C). Zum vierten die Umsetzung des Konzepts in ersten gesamthaften Pilot-Anwendungen inklusive der Auslotung der Verfahrensleistungen in neuen Inhaltsgebieten (D).

(A) Der NMDS-Algorithmus RobuScal erlaubt die Interpretation von NMDS-Lösungen auf individueller Ebene. Mathematische Korrekturverfahren erlauben es, stabile und gegen Ausreißer in den Daten robuste Objektverhältnisse darzustellen (Läge, Daub, Bosia, Jäger & Ryf, 2005). Ungeklärt bleibt die Frage, ob wirklich ein Zusammenhang zwischen der Kartenqualität und der über andere Herangehensweisen erhobene Wissensabschätzung (z.B. Multiple-Choice) besteht (externe Validierung), ob das Verfahren also das misst, was es zu messen vorgibt. Des Weiteren muss gezeigt werden, dass Wissensveränderungen detektiert werden können. Im Verlauf eines merkmalsbasierten Lernprozesses gewinnt das Wissen an Umfang und an Qualität. Diese Gerichtetheit im Lernverlauf im Sinne einer strukturellen Annäherung an ein Kriterium muss für die Bestätigung der Verfahrensgüte gegeben sein, denn es besteht die Möglichkeit, dass zwar Merkmalswissen aufgenommen, dieses sich jedoch ohne Lernunterstützung nicht merklich in der relationalen Ordnung des Lernerwissens auswirkt. Diese beiden Fragen werden im Kapitel 2 „Die Veränderung von Wissenskarten beim Merkmalslernen“ behandelt. Das Kapitel 3 „Der NMDS-Stress als Maß für Assimilation und Akkomodation beim Lernen von Wissensstrukturen“ ist als Exkurs zu verstehen. Es soll gezeigt werden, dass die NMDS auch sehr feine und meist unterschwellige, methodisch nur schwerlich fassbare kognitive Konzepte zu beobachten in der Lage ist wie Assimilations- und Akkomodationsprozesse (Piaget, 1976) beim Lernen.

² Die genannten Arbeiten wurden im Rahmen eines Projekts des Schweizerischen Nationalfonds unter der Leitung von Prof. Dr. Damian Läge, Universität Zürich, Psychologisches Institut, Angewandte Kognitionspsychologie mit dem Namen Merkmalsorientierte Diagnostik in adaptiven Lernsystemen (MODAL) durchgeführt (Projektnummer 1114-63536.00).

(B) Der zentrale Punkt eines adaptiven, Intelligenten Tutoriellen Systems ist das Erkennen von Wissensmängeln. Werden Sie nicht entdeckt, ist keine inhaltliche Adaptivität des Verfahrens möglich und würde sich somit nicht von einem einfach gestrickten, selbstgesteuerten Lernprozess unterscheiden. Das Kapitel 4 „*Der ObjLoss einer Prokrustes-Transformation als Maß für Wissen*“ beinhaltet ein Experiment und eine Simulation, in denen überprüft wird, unter welchen Voraussetzungen die NMDS und die Prokrustes-Transformation nicht- oder falsch-gewusste Objekte einer Person detektieren kann, wie groß die Vorhersagegenauigkeit ist und gibt darauf aufbauend Empfehlungen für einen automatisierten Praxiseinsatz.

(C) Globale Ähnlichkeitseinschätzungen über einen Satz von Objekten bilden die Datengrundlage zur Messung des relationalen Merkmalswissens. Klassische NMDS-Verfahren benötigen zur Bildung einer Wissenskarte einen vollständigen Satz an paarweisen Vergleichen zwischen allen Objekten. Mit steigender Objektmenge des Wissensgebietes vergrößert sich die benötigte Urteilsmenge exponentiell. Sehr schnell wird somit die Grenze der in der Praxis zumutbaren Urteilsmenge überschritten. Das Kapitel 5 „*Effizienzsteigerung in der ähnlichkeitsbasierten Wissensdiagnostik*“ überprüft ein schlankeres, sequentielles Diagnostikprozedere auf seine Trennschärfe hin und erweitert beziehungsweise spezifiziert daraus folgernd das (ehemalige) didaktische Konzept von INTUS (vgl. Abbildung 1) unter Berücksichtigung lerntheoretischer Modelle und methodischer Anforderungen.

(D) Nachdem bislang grundlegende Verfahrensleistungen in kontrollierten Laborstudien überprüft wurden, handelt es sich in den weiteren Ausführungen zum einen um die Umsetzung der gewonnenen Erkenntnisse in einen größeren Zusammenhang, zum anderen um die Überprüfung des Verfahrens in realen Settings und neuen Inhaltsgebieten.

Das Kapitel 6 „*Adaptive Lernempfehlungen über Prokrustes-Transformation und Kohonennetze*“ stellt eine Simulation und ein Experiment vor, welche das herkömmliche Verfahren der Wissensabschätzung mittels NMDS und Prokrustes-Transformation auszuweiten versucht. Die Prokrustes-Transformation erlaubt lediglich die Qualitätsbeurteilung auf Objektebene. Kohonennetze sind selbstorganisierende Merkmalslandschaften und bieten die Möglichkeit, noch eine Ebene tiefer zu gehen und anhand von relationalen Urteilen das Merkmalswissen einer Person zu beurteilen. Es wird das adaptive INTUS-Konzept anhand von künstlichen Lernern unter verschiedenen Bedingungen getestet und erweitert eine Pilotsimulation von Daub (2001) konzeptuell. Ein empirisches und gleich wie die Simulation strukturiertes Lernexperiment versucht, die aus der Simulation gewonnenen Erkenntnisse im realen Umfeld zu bestätigen.

In den bisherigen, stark kontrollierten Lernexperimenten kamen zwei prototypische Inhaltsgebiete zum Einsatz. Zum einen das komplexer strukturierte Fachgebiet *Ornithologie* und zum anderen das zur Überprüfung für maximale Verfahrensleistungen herangezogene Wissensfeld der *Geographie (Länder)*. Kapitel 7 „*A Structure-Based Expert Model of the ICD-10 Mental Disorders*“ beschäftigt sich mit der Frage, ob psychopathologische Störungsbilder als neues Inhaltsgebiet ebenfalls mit der NMDS skalierbar sind. Befragungen von Experten aus der Psychiatrie und Klinischer Psychologie lieferten die Testdaten (Expertenmodell). Das damit zusammenhängende Kapitel 8 „*Das Zürcher Wissensstrukturmodell für Psychopatho-*

logie (ZWP) – eine Pilotstudie“ verwendet das vorher beschriebene Expertenmodell als Kriterium zur Beurteilung des Wissensstandes von Teilnehmern eines postgradualen Weiterbildungskurses. Ziel ist es, die Validität der NMDS-Methodik inklusive Prokrustes-Transformation in einem realen curricularen Umfeld mit dem Ziel des Einsatzes der Verfahrenslogik als komplementäres Qualitätssicherungsinstrument zu bestätigen (als Pendant zu den Laborstudien in Kapitel 2 „*Veränderung von Wissenskarten beim Merkmalslernen*“).

Kapitel 9 „*Pilot-INTUS*“ schließlich stellt ein umfassendes Lernexperiment vor, welches alle bislang generierten Erkenntnisse in einer Großstudie zu replizieren versucht. Laien in Ornithologie werden schrittweise erst zu einem Grundwissen herangeführt, welches anschließend assimilativ erweitert wird. Die Hauptfragestellungen aus Kapitel 2 und 4 sollen repliziert sowie das reduzierte Diagnostikkonzept aus Kapitel 5 erstmals in einer praxisnäheren Anwendung eingesetzt werden.

1.4. Gesamtdiskussion der Resultate

In der Folge werden die Ergebnisse aus den vier eben beschriebenen Schwerpunkten diskutiert und summarisch zusammengefasst. Ausführliche Interpretationen der Resultate sind in den jeweiligen Diskussionsabschnitten der einzelnen Kapitel zu finden.

(A) In zwei Lern-Experimenten und unterschiedlich komplexen Inhaltsgebieten konnte der Nachweis erbracht werden, dass alleine aufgrund des Lernens von Merkmalsinformation sich die kognitiven Wissensstrukturen den neuen Gegebenheiten anpassen. Die strukturbildende Kraft des Gedächtnisses lässt sich anhand Kognitiver Karten (NMDS) nachzeichnen, indem festgestellt werden kann, dass sich über einen Lernverlauf die individuellen Wissenskarten gerichtet einem Expertenmodell annähern. Außerdem verringert sich die Idiosynkratie der individuellen relationalen Wissensgebilde und erreicht zum Schluss eine ansprechende Homogenität im Vergleich zu anderen Personen, die das gleiche Lernmaterial bearbeitet haben. Eine weitere wichtige Erkenntnis aus der Studie „*Die Veränderung von Wissenskarten beim Merkmalslernen*“ ist, dass die NMDS wirklich die Umsetzung von Merkmalswissen in relationale Gebilde messen kann aufgrund des signifikanten Zusammenhangs zwischen der Kartengüte und der gespeicherten Merkmalsmenge, gemessen über einen Multiple-Choice-Fragebogen. Damit ist eine wichtige Grundlage gegeben, NMDS-Lösungen überhaupt als Maß für den aktuellen Wissensstand und der Beobachtung von Lernfortschritten einer Person zu verwenden.

Eine Nebenauswertung der gleichen Experimente führte zu einem interessanten Befund. Die seit längerem bekannten und wissenschaftlich anerkannten theoretischen Lernkonzepte der Assimilation und Akkomodation wurden – unseres Wissens – bislang noch nie in der vorliegenden Weise empirisch überprüft. Lediglich das Endresultat eines Lernvorgangs wurde in der Weise interpretiert, ob eine Person Wissen assimiliert hat oder aufgrund der neuen Information eine Akkomodationsleistung notwendig war. Anhand des NMDS-Stress – einem primär technischen Maß – konnten die zeitlich getrennt ablaufenden Prozesse einer exakteren Überprüfung unterzogen werden. Ausgehend von hypothetischen Vorhersagen zeigte sich die NMDS sensitiv genug, die assimilative Neuaufnahme von Wissen anhand erhöhter Inkon-

sistenzen in den Ähnlichkeitsurteilen nachzuweisen, gefolgt von einer Erhöhung der Urteilkonsistenz in der Zeit zwischen den Lernphasen, in denen sich das Wissen „setzen“ konnte. Dies ist ein eindrücklicher Befund, der Hinweise auf die breit gefächerten Anwendungsmöglichkeiten der NMDS-Methodik gibt.

(B) Das Kapitel 4 „*Der ObjLoss einer Prokrustes-Transformation als Maß für Wissen*“ spezifiziert die Messleistung der NMDS, indem die nächst tiefere Ebene der abhängigen Variable „Prokrustes-Transformation“ unter die Lupe genommen wird. War es unter Punkt (A) noch die mittlere Kartenabweichung (AvgLoss), wird hier der Fokus auf die einzelnen Objektabweichungen (ObjLoss) gelegt. Ein kontrolliertes Lern-Experiment diente der Bestimmung der Trefferrate für mangelhaft gewusste Objektinformation. Unter der Voraussetzung eines bestimmten Wissensniveaus (rund 80% korrekt) gibt die Prokrustes-Transformation mit geringem Fehleranteil diejenigen Objekte als mangelhaft gewusst zurück, die in Wirklichkeit auch nicht gelernt worden sind. Eine Simulation mit dem im Vergleich zur Ornithologie einfacher strukturierten Inhaltsgebiet *Länderwissen* lotete anschließend die maximale Verfahrensleistung aus. Zwei wichtige Erkenntnisse sind daraus entstanden: Zum einen, dass die zu skalierende Objektmenge um die 20 Einheiten umfassen muss, um eine ausreichende Stabilität in der Treffergenauigkeit zu erreichen (vgl. zur rein statistisch-mathematischen, für robuste Lösungen notwendigen Menge von 15 Objekten Läge, Daub, Bosia, Jäger & Ryf, 2005). Zum anderen zeigte sich, in welchem Bereich die NMDS (bzw. der ObjLoss) ihre maximale Performanz aufweist. Erst ab einem Vorwissensniveau von um die 60% korrekt gewusst, kann ohne größere Falsch-Alarm-Rate mangelhaftes Wissen detektiert werden. Diese Resultate stimmen zuversichtlich, anhand von Wissenskarten einzelne schlecht repräsentierte Informationseinheiten zu detektieren und dem adaptiven Prozess zu übergeben. Daraus folgt auch, dass Wissenskarten nicht in der Lage sind, den gesamten Lernverlauf, angefangen von „Null-Wissen“, diagnostisch zu begleiten. Hilfreich sind sie in der Feinregulierung von spezifischen Defiziten, wobei diese eingeschränkte Anwendungsmöglichkeit implizit natürlich bereits im Adaptivitätsgedanken enthalten ist. Mit anderen Worten: Was nützt es, adaptiv zu sein, wenn der Lerner noch nicht mal über grundlegendes Basiswissen verfügt?

(C) Ein großer Nachteil der Skalierungsmethodik ist die hohe Menge an benötigten Rohdaten. Die traditionelle Erhebung paarweiser Ähnlichkeitsbeurteilungen zwischen allen interessierenden Objekten führt bereits bei einer relativ geringen Objektzahl zu einem in der Praxis nicht mehr zumutbaren Diagnostikaufwand. Ein gestuftes Diagnostikkonzept entschärft diese Problematik dahingehend, dass nicht mehr ein exponentieller Zusammenhang zwischen Objektzahl und Urteilsmenge, sondern lediglich ein linearer bestehen bleibt. Wenn Probanden als Erstes eine Grundmenge lernen, welche immer noch vollständig beurteilt werden muss, reicht es aus, neu gelerntes Wissen (Objekte) nur noch mit den Objekten des Vorwissens zu beurteilen, ohne dass sich die NMDS-Lösungen maßgeblich von einer so genannten Vollerhebung unterscheiden. Diese Erkenntnis bietet erst die Möglichkeit, das theoretische Konzept ökonomisch für die Praxis umzusetzen und eine effiziente, aber dennoch umfassend vollständige Wissensbeurteilung durchzuführen. Aufgrund dieser Resultate und theoretischer Überlegungen passt sich auch das didaktische Konzept einem gestuften Prozess an. Vorab wird dem Lerner für das Inhaltsgebiet repräsentatives Basiswissen vermittelt. Darauf aufbauend wird

der Wissensraum assimilativ erweitert. Dieses Vorgehen wird in der pädagogischen und didaktisch orientierten Literatur als ein optimaler Weg beschrieben (z.B. Ausubel, 2000; Steiner, 1996), womit die auf den ersten Blick als methodische Einschränkungen erscheinenden Erkenntnisse in optimaler Weise lerntheoretische Paradigmen ergänzen.

(D) Daub (2001) zeigte in einer Simulation sehr eindrücklich, wie sich die adaptive Auswahl von mangelhaft gewussten Elementen positiv auf den Lernprozess auswirkt. Die Lernempfehlungen basierten nicht nur auf den Ergebnissen einer Prokrustes-Transformation, sondern ein neuronales Netz (Kohonen-Netz) erweitert die ursprüngliche Diagnostikidee mit dem Ziel, einzelne mangelhaft gespeicherte Merkmalsausprägungen zu detektieren. Im Vergleich zu einem Lerner, welche in zufälliger Reihenfolge alle Informationen einmal lernt, ist der adaptive Lerner viel schneller, das heißt, mit weniger Lernaufwand auf einem Expertenniveau. Ein möglicher Kritikpunkt an diesem Vorgehen war das Nicht-Berücksichtigen von Lern- und Vergessensraten. Ein aus diesem Grund durchgeführtes, empirisches Lern-Experiment ergab leider aus experimentaltechnischen Mängeln keine nennenswerten Erkenntnisse. Das komplex strukturierte Lernsetting ermöglichte es den Probanden nicht, in der vorgegebenen Zeit eine adäquate Wissensstruktur zu bilden. Zu betonen ist, dass dies nicht grundsätzlich gegen die gestellten Hypothesen spricht, da bereits bestimmte notwendige Voraussetzungen wie z.B. ein bestimmtes Vorwissensniveau nicht gegeben waren. Erneute Forschung wäre notwendig, den Adaptivitätsgedanken an sich und die Tauglichkeit eines unterstützenden Kohonen-Netzes in der Empirie zu bestätigen. Es muss trotzdem in Frage gestellt werden, ob aufgrund der vielen notwendigen Voraussetzungen ein solcher Ansatz in der Praxis umsetzbar ist. Das um die Faktoren Lernen und Vergessen erweiterte Simulations-Konzept von Daub (2001) stellt ein Verfahren vor, in welchem die Adaptivitätslogik aufgrund der schwierig umsetzbaren Vorbedingungen nochmals überprüft werden soll. Es ist zu erwarten, dass der adaptive Lerner immer noch im Vorteil hinsichtlich der Lerneffizienz steht. Dies auch noch, wenn der Vergleichslerner die wichtige Chance erhält, mehrfach Merkmalsätze zu lernen (auf die Praxis übertragen kommt dies einem selbstgesteuerten Kontrollprozess gleich.).

Die breite Anwendungsmöglichkeit der NMDS in verschiedenen Inhaltsgebieten ergibt sich aus den Resultaten in den Kapiteln 7 „*A Structure-Based Expert Model of the ICD-10 Mental Disorders*“ und 8 „*Das Zürcher Wissensstrukturmodell für Psychopathologie (ZWP) – eine Pilotstudie*“. Expertenbefragungen haben ergeben, dass trotz der in der Literatur beschriebenen, kategorialen Struktur von psychopathologischen Störungsbildern (z.B. das Klassifikationssystem ICD-10) die individuellen Expertenmeinungen nicht dieser Logik entsprechen. Vielmehr scheint die NMDS geeignet, dieses Inhaltgebiet realitätsnäher zu messen. Eine Pilotstudie mit Weiterzubildenden bestätigte dann erstmals in einem realen Umfeld, dass NMDS und Prokrustes-Transformation als Evaluationsinstrument zur Sicherstellung des Lernfortschritts einsetzbar sind.

Kapitel 9 „*Pilot-INTUS*“ stellt ein Experiment vor, welches empirisch all die gewonnenen Erkenntnisse aus den Einzelstudien in einem Gesamtzusammenhang umzusetzen versucht. Erstmals wird das gestufte didaktische und methodische Konzept in vollem Umfang angewendet und sollte einerseits als Existenznachweis für die Tauglichkeit des Systems und andererseits der Replikation bestehender Resultate dienen. Die Resultate zeigen nur sehr bedingt

in die vorhergesagte Richtung und das Experiment muss insgesamt als gescheitert betrachtet werden. Das enorm komplexe Experimentalsetting überforderte die Möglichkeiten der sehr wenigen Lerner, in der vorhandenen Zeit die zu erwartenden Lern- und Strukturierungsleistungen zu erbringen. Nichtsdestotrotz zeigen die Resultate in die vorhergesagte Richtung, können aber nicht als genügend valide betrachtet werden.

Die valide und effiziente Diagnostik von merkmals- und objektbasiertem Sachwissen steht im Forschungs-Zentrum des Intelligenten Tutoriellen Systems INTUS, welches an der Universität Zürich, Psychologisches Institut, Angewandte Kognitionspsychologie entwickelt wird. Mit der Methode der Nonmetrischen Multidimensionalen Skalierung (NDMS) können strukturell-relationale Ordnungen des subjektiven Faktenwissens visualisiert und über die Prokrustes-Transformation auf ihre Qualität hin beurteilt werden. Ein robuster Skalierungsalgorithmus sowie eine robuste Prokrustes-Transformation bilden das diagnostische Methodeninventar zur Erkennung von individuellen Wissensmängeln. Eine ausreichende Diskriminationsleistung der Verfahrenslogik vorausgesetzt, ergibt sich die Möglichkeit zur Entwicklung eines sich adaptiv an den Wissensstand anpassenden Lernsystems, indem es der Person diejenigen Informationen empfiehlt, welche als mangelhaft gewusst klassifiziert werden.

Die Studien haben gezeigt, dass die relationale Wissensdiagnostik gewissen Einschränkungen unterliegt, etwa ein notwendiges, bestimmtes Maß an Vorwissen für eine genügend hohe Vorhersagegenauigkeit oder eine bestimmte Menge an Objekten für eine stabile NMDS-Lösung. Trotzdem stimmen die Resultate zuversichtlich, über ein gestuftes, modulares didaktisches Konzept, welches zum einen zum Zweck einer effizienteren Wissensdiagnostik und zum anderen aufgrund lerntheoretischer Empfehlungen an einen optimal strukturierten Lernprozess, Adaptivität der vom System empfohlenen Lerninhalte zu gewährleisten.

Anwendung kann diese Verfahrenslogik in computerbasierten Lernsystemen finden (Intelligente Tutorielle Systeme ITS, E-Learning, Computer Based Training CBT usw.). Die weit gefächerten Einsatzmöglichkeiten erlauben einen flexiblen Umgang mit dem Verfahren. Nicht nur die Verwendung als diagnostisches Instrument, sondern ebenfalls als mögliche didaktische Lernunterstützungshilfe bei der Erkennung relationaler Beziehungen fördert den Lernprozess in Richtung einer hoch elaborierten und für Wissenstransfers und -umsetzungen notwendigen, differenziert geordneten kognitiven Wissensstruktur.

1.5. Literatur

- Albert, D. & Lukas, J. (Eds.). (1999). *Knowledge spaces. Theories, empirical research and applications*. Mahwah: Erlbaum.
- Ausubel, D. P. (2000). *The acquisition and retention of knowledge*. Dodrecht: Kluwer Academic Publishers.
- Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling – Theory and applications*. New York: Springer.
- Charney, D., Reder, L. & Kusbit, G. W. (1990). Goal setting and procedure selection in acquiring computer skills. A comparison of tutorials, problem solving and learner exploration. *Cognition and Instruction*, 7, 323-342.
- Crowder, N. A. (1959). Automating tutoring by means of intrinsic programming. In E. Galanter (Ed.), *Automatic teaching. The state of the art* (pp. 109-116). New York: John Wiley and Sons, Inc.

- Daub, S. (2001). *Similarity Simulation – Wie man den Code des Globalurteils knackt*. Dissertationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Goldstein, I. P. (1982). The genetic graph: A representation for the evolution of procedural knowledge. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutorial systems* (pp. 51-78). New York: Academic Press, Inc.
- Gower, J. C. & Dijksterhuis, G. B. (2004). *Procrustes problems*. Oxford: Oxford University Press.
- Klein, B. (2000). *Didaktisches Design hypermedialer Lernumgebungen. Die adaptive Lernumgebung „incops“ zur Einführung in die Kognitionspsychologie*. Marburg: Tectum.
- Läge, D. (2001). *Ähnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Läge, D., Daub, S. Bosia, L., Jäger, C. & Ryf, S. (2005). *Die Behandlung ausreißer-behafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag*. AKZ-Forschungsbericht Nr. 21. Zürich: Angewandte Kognitionspsychologie.
- Leutner, D. (1992a). *Adaptive Lernsysteme. Instruktionspsychologische Grundlagen und experimentelle Analysen*. Weinheim: Psychologie Verlags Union.
- Leutner, D. (1992b). Das Testlängendilemma in der lernprozess-begleitenden Wissensdiagnostik. *Zeitschrift für Pädagogische Psychologie*, 24, 232-248.
- Leutner, D. (1993). Das gleitende Testfenster als Lösung des Testlängendilemmas: Eine Robustheitsstudie. *Zeitschrift für Pädagogische Psychologie*, 7, 33-45.
- Leutner, D. (1995). Adaptivität und Adaptierbarkeit multimedialer Lehr- und Informationssysteme. In L. J. Ising & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia* (S. 140-149). Weinheim: Psychologie Verlags Union.
- Mandl, H. & Fischer, F. (Hrsg.). (2000). *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*. Göttingen: Hogrefe.
- Piaget, J. (1976). *Die Äquilibration der kognitiven Strukturen*. Stuttgart: Klett.
- Pressey, S. L. (1926). A simple apparatus which gives tests and scores and teaches. In (reprint) A. A. Lumsdaine & R. Glaser (Eds.). (1960), *Teaching machines and programmed learning* (pp. 35-41). Washington D.C.: National Education Association of the United States.
- Roskam, E. E. & Lingeos, J. C. (1970). MINISSA-I: A Fortran IV program for the smallest space analysis of square symmetric matrices. *Behavioral Science*, 15, 204-205.
- Siemer, J. & Angelides, M.-C. (1998). Towards an intelligent tutoring system architecture that supports remedial tutoring. *Artificial Intelligence Review*, 12, 469-511.
- Skinner, B. F. (1958). Teaching machines. In (reprint) A. A. Lumsdaine & R. Glaser (Eds.). (1960), *Teaching machines and programmed learning* (pp. 137-158). Washington D.C.: National Education Association of the United States.
- Steiner, G. (1996). *Lernen – Zwanzig Szenarien aus dem Alltag* (2. vollst. überarb. Aufl.). Bern: Huber.
- Strohschneider, S. (1990). *Wissenserwerb und Handlungsregulation*. Wiesbaden: Deutscher Universitätsverlag.
- Van Merriënboer, J. J. G., Van Den Berg, K. G. & Maaswinkel, D. M. (1989). Some experiences with two intelligent tutoring systems for teaching computer programming: PROUST and LISP-tutor. In J. M. Pieters (Ed.), *Intelligent tutorial systems and instruction* (pp.7-44). Enschede: OTG Onderwijsleerprocessen.
- Young, F. W., & Lewyckyj, R. (1979). *ALSCAL-4: Users guide*. Chapel Hill: Data Analysis and Theory Associates.

2.

**Die Veränderung von
Wissenskarten beim
Merkmalslernen**

Zusammenfassung

Merkmalswissen über eine Anzahl Objekte lässt sich relational in Form einer Wissenskarte abbilden. Mit fortschreitendem Lernen sollten sich solche individuellen Karten systematisch verbessern. Dazu wird folgendes Messverfahren vorgeschlagen: Eine Matrix paarweiser Ähnlichkeitsurteile zwischen einem Satz von Objekten wird mittels Nonmetrischer Multidimensionaler Skalierung in eine zweidimensionale Wissenskarte überführt. Diese Karte wird durch Prokrustes-Transformation auf eine Zielstruktur gelegt. Die numerische Abweichung der beiden Karten (der AvgLoss) dient dann als kritisches Maß für die Güte des Wissens.

Zwei Lernexperimente zeigen den angenommenen Zusammenhang zwischen Merkmalswissen und Güte der Wissenskarte auf. Experiment 1 überprüft das vorgeschlagene Verfahren mittels eines komplexen Wissensgebietes (der Ornithologie) und belegt die systematische Verbesserung der Karten. Experiment 2 dokumentiert dann an einfacherem Lernmaterial (Länder) die Konvergenz der Wissenskarten untereinander und gibt einen ersten Hinweis darauf, wie ähnlich individuelle Wissenskarten einer Zielstruktur werden können.

Schlüsselwörter: Wissenspsychologie, Sachwissen, Merkmalswissen, Wissenskarte, Kognitive Struktur, Kognitive Karte, Lernexperiment, Ähnlichkeitsurteil, Nonmetrische Multidimensionale Skalierung, Prokrustes-Transformation

2.1. Einleitung

Elemente unseres Wissens existieren nicht einfach unverbunden nebeneinander, sondern sie bilden zusammen ein Ganzes. Diese alte Erkenntnis der Piaget'schen Tradition in der Pädagogischen Psychologie (Aebli 1976; Ausubel, 2000; Inhelder, 1974; Inhelder, Sinclair & Bovet, 1974; Montada, 1995; Piaget, 1973, 1976, 2001; Piaget & Inhelder, De Caprona & Cornu-Wells, 1987; Steiner, 1996) hat die Wissenspsychologie seit ihrer Wiederbegründung als eigenständiger Forschungszweig (vgl. Mandl & Spada, 1988) in ihren neueren Entwicklungen aufgenommen. Die modernen Modelle der Wissensmodellierung zeichnen sich durch eine Art System aus, d.h. die Elemente konstituieren unwillkürlich einen Satz von Relationen. Z.B. unter Rückgriff auf Konzeptionen eines Semantischen Netzwerkes (wie das von Quillian, 1968) wurde diese relationale Facette von Sachwissen aufgegriffen und mit Mapping-Techniken abgebildet. Damit tragen diese Modelle dem Umstand Rechnung, dass sich Wissensstrukturen durch Differenziertheit, Vollständigkeit und Komplexität auszeichnen, also weit über eine einfache Anhäufung unverbundener Fakten hinausgehen (vgl. Mandl & Fischer 2000). Dabei bedienen sich die Mapping-Techniken vorwiegend graphischen Verfahren in der Datenerhebung. Exemplarisch zu nennen sind die zahlreichen Arbeiten mit der Heidelberger Strukturlegetechnik (basierend auf Scheele & Groeben, 1984) und die verwandten Verfahren der graphentheoretischen Darstellung von Wissensstrukturen (Bonato, 1990), die Netzwerk-Elaborierungs-Technik (Eckert, 1998, 1999, 2000), das Concept Mapping Software Tool (COMASOTO) zur Diagnose strukturellen Wissens (Weber & Schumann, 2000), die Wis-

sensvermittlung, tutoriell unterstützte Wissensanwendung und Wissensdiagnose mit Begriffsnetzen nach Plötzner, Beller & Härder (2000) oder die Knowledge-Tracking-Methode von Janetzko & Strube (2000).

Mandl & Fischer (2000) resümieren, dass nach anfänglichen Problemen bei der Wissensmodellierung (hoher Idiosynkrasie-Anteil, fragliche Validität und mangelnde Re-Test-Reliabilität) inzwischen theoretische Grundlagen und auch konkret eingesetzte Diagnoseverfahren und Benutzerinterfaces vorliegen. Sie verweisen in ihrem historischen Abriss auch auf die psychometrisch-strukturelle Perspektive, die „... mit verschiedenen Formen der Ähnlichkeitsskalierungen und ihrer graphischen Darstellung (z.B. Multidimensionale Skalierung, Clusteranalysen) ... [den Versuch unternehmen] ... psychische Phänomene und deren Veränderung mathematisch zu quantifizieren.“ (S. 4). Über die vorhandenen graphischen Mapping-Techniken hinaus sprechen die Autoren auch explizit die Möglichkeit an, auf der Basis nicht-graphischer Daten (z.B. multiple Paarvergleiche) über komplexe Algorithmen auf kognitive Wissensstrukturen zu schließen.

Im Folgenden wird ein Weg zur Diagnose merkmalsbezogenen Sachwissens vorgestellt, der genau einer solchen Grundidee folgt: Über numerische relationale Urteile soll für einen Objektbereich festgestellt werden, wie eine individuelle Person dieses Feld für sich selbst strukturiert hat. Ein geeignetes, weil verhältnismäßig expertiseunabhängiges Messinstrument dafür sind paarweise Ähnlichkeitsurteile (vgl. Klauer, 1989). Bereits mit wenigen Informationen sind wir Menschen nämlich in der Lage, ein ungefähres Urteil über die Ähnlichkeit zweier Dinge oder Personen abzugeben. Dabei vertrauen wir offenbar auf unser „gutes Auge“, mit dem wir wichtige von unwichtigen Merkmalen unterscheiden und die verbleibenden wichtigen Merkmale in eine wohlproportionierte Relation zueinander bringen. Laien sind genauso in der Lage wie Experten, die Ähnlichkeit zwischen zwei Dingen abzuschätzen – auch wenn das Urteil natürlich beim Experten qualitativ besser, d.h. sachadäquater im Vergleich zu einer Norm ausfallen sollte. Läge (2004) konnte zeigen, dass sich Ähnlichkeitsurteile mit zunehmendem Expertisegewinn systematisch verändern.

Nach unserem Dafürhalten teilen Ähnlichkeitsurteile mit den bestehenden Mapping-Techniken gegenüber einer reinen Abfrageliste von Sachfragen den Vorteil, dass sie relationale Gefüge, Integration und Elaboration (Gewichtungen, kategoriale Strukturierung usw.) der Wissensstruktur mit messen. Zusätzlich sind sie unabhängig von Frageformaten und relationalen Vorschriften. Allerdings liegt die resultierende relationale Information bei Ähnlichkeitsurteilen genauso wenig unmittelbar auf der Hand wie das zu ermittelnde Merkmalswissen über die einzelnen Objekte (welches mit Mapping-Techniken nur umständlich zu erheben ist). Es bedarf vielmehr komplexer Auswertalgorithmen, um relationales Wissen und Merkmalswissen aufzuzeigen.

Durch Ähnlichkeitsurteile ermittelte Proximitäten zwischen Paaren von Objekten lassen sich mittels Nonmetrischer Multidimensionaler Skalierung (NMDS) als Distanzwerte verrechnen und zur Grundlage von „Wissenskarten“ machen (die Logik dieses Verfahrens ist anwendungsorientiert bei Borg & Groenen, 1997 und mathematisch grundlegender bei Mathar, 1997 beschrieben). Auf diese Weise gewinnt man die relationale Information über das aktuelle

Sachwissen. Solche individuellen „Landkarten“ können durch Prokrustes-Transformation mit einem Expertenmodell verglichen werden. Bereits ohne großen mathematischen Aufwand springt bei einem solchen Vergleich ins Auge, welche Objekte ein Lerner gut kennt (d.h. korrekt verortet) und welche er falsch platziert (als Beispiel vgl. Abbildung 2).

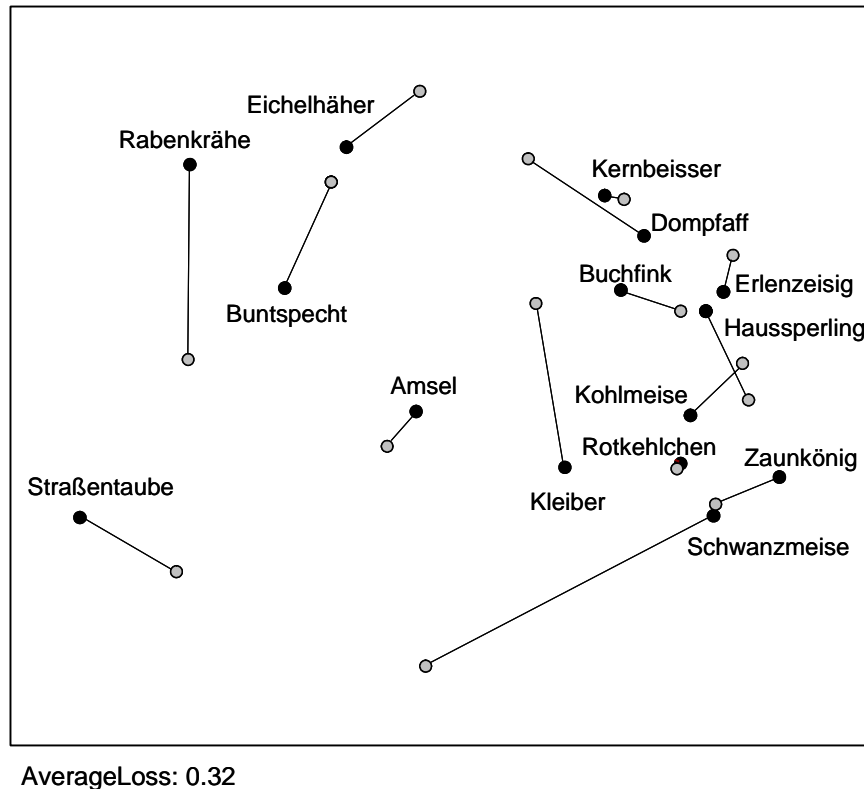


Abbildung 2. Prokrustes-Transformation eines Expertenmodells mit einer Lernerkarte. Die schwarzen Punkte stehen für die Positionen der Vogelarten im Expertenmodell, die grauen für diejenigen in der Lernerkarte (vgl. Experiment 1).

Als Gesamtabweichungswert einer Prokrustes-Transformation hat Läge (2001) den Average-Loss (kurz: AvgLoss) eingeführt. Dieses Pauschalmaß könnte also als Schätzung für die Güte des aktuellen Wissensstandes einer Person herangezogen werden. Im Laufe des Expertisegewinns sollte sich dieser Abweichungswert von einer Wissensnorm systematisch verringern.

Ein solcher Vorschlag, um die Güte individuellen Wissens in einer vernetzten Form zu berechnen, ist unseres Wissens bislang noch nicht gemacht worden. Zwar gibt es eine Reihe von Anwendungen Nonmetrischer Multidimensionaler Skalierung in der Psychologie (etwa Gigerenzer, 1981; Marx & Hejj, 1989; Marx & Läge, 1995; Edelman, 1998; Lee, Byatt & Rhodes, 2000), und auch Prokrustes-Transformationen wurden vereinzelt bereits durchgeführt (vgl. Gower & Dijksterhuis, 2004), aber geschah dieses doch durchgängig für gemittelte Proximitätsmatrizen befragter Personenstichproben. Technisch betrachtet hatte das seinen guten Grund, denn die bislang bestehenden Verfahren der Multidimensionalen Skalierung waren sehr anfällig für Ausreißerwerte in einer Proximitätsmatrix. Solche Ausreißer (in Form extremer Inkonsistenzen hinsichtlich des Gesamt der Messwerte) passieren immer einmal wieder bei der Abgabe von Urteilen, sind jedoch a priori nicht zu detektieren. Jedoch können sie die resultierenden NMDS-Karten in der Gesamtstruktur massiv beeinflussen, indem alle

Punkte in ihrer Struktur verändert werden, nur um einen einzelnen zu großen Einpassungsfehler abzufangen. Bei gemittelten Stichprobenwerten verschwinden solche unsystematischen Ausreißer, weswegen sie dann für die Skalierung kein Problem mehr darstellen. Erst mit der Entwicklung eines robusten Verfahrens der NMDS, welches Ausreißer im Prozess der schrittweisen Genese der geometrischen Konfiguration detektiert und eliminiert, lassen sich nun aber auch Wissenskarten auf individuellem Niveau verlässlich nachzeichnen (RobuScal-Algorithmus; Läge, Daub, Bosia, Jäger, & Ryf, 2005). Analoges gilt für die Prokrustes-Transformation: Auch hier kann ein einzelnes schlecht verortetes Objekt das Übereinanderlegen der Strukturen massiv beeinflussen. Deswegen ist auch hier eine robuste Version angebracht (Läge, 2001).

Dank dieser robusten Verfahren ist es nun möglich, die individuellen Ähnlichkeitsurteile in Wissenskarten umzusetzen. Dabei ist es interessant, die Veränderung dieser Wissenskarten im Verlauf der Beschäftigung mit den Merkmalen von Objekten zu beobachten. Die Zunahme oder (im Falle inadäquaten Vorwissens) die Veränderung von Merkmalswissen sollte ja eine Auswirkung auf die relationalen Ähnlichkeitsurteile haben, die eine Person abgibt. Und die Matrix dieser veränderten Ähnlichkeitsurteile wiederum sollte eine Restrukturierung der Karte als ganzer oder zumindest einzelner Objekte innerhalb einer als ganzer stabil bleibenden Karte bewirken.

Dabei setzt diese Kette eine gewisse Integration des durch die Beschäftigung neu erworbenen Wissens voraus. Merkmale, die ohne jedes Verständnis einfach nur auswendig gelernt werden, mögen in reproduktiven Testfragen (wie z.B. den meisten Multiple-Choice-Aufgaben) abrufbar sein. In die Veränderung eines Ähnlichkeitsurteils sollten sie aber nur dann eingehen, wenn die tatsächliche Bedeutung für dieses Objekt und für das Gesamt des Wissensfeldes auch verstanden wurde.

Im Folgenden werden zwei systematische Lernstudien vorgestellt, die sich der Frage der Veränderung von Wissenskarten beim Merkmalslernen widmen. Experiment 1 bedient sich dazu des Wissensgebietes der Ornithologie, weil hier hinreichend klar definierte Objekte (Vogelarten) mit komplexen definierenden und charakterisierenden Merkmalssätzen vorliegen. Experiment 2 schließt im Aufbau an Experiment 1 an, versucht aber den Gedanken der Konvergenz der individuellen Wissenskarten noch stärker herauszuarbeiten. Dazu wird dem Lernexperiment der Inhaltsbereich der Geographie zugrunde gelegt, und zwar sind Länder anhand von drei klar definierten Merkmalen (Bruttosozialprodukt, Einwohnerzahl und Größe) auf ihre Ähnlichkeit hin einzuschätzen. Mit diesem – im Vergleich zu Experiment 1 leicht lern- und integrierbaren – Material kann der maximale Grad an Konvergenz individueller Wissenskarten ausgelotet werden. Am Ende dieser beiden Experimente soll dann ein Einblick in Ausmaß und Richtung der Veränderung einer Wissenskarte beim Lernen stehen.

2.2. Experiment 1

2.2.1. Fragestellung und Hypothesen

Experiment 1 misst den Fortschritt des Lernens in individuellen Wissenskarten, indem es den jeweiligen Kenntnisstand in einem konvergenten Wissensgebiet (der Ornithologie) mit einem Expertenmodell vergleicht. Dabei wird als erster Hauptbefund eine gerichtete und zudem konvergente Veränderung der individuellen Wissenskarten untereinander erwartet. Zweitens sollte diese Veränderung in eine systematische Annäherung der individuellen Wissenskarten an ein Expertenmodell resultieren (Im Idealfall sollte ein Lerner so dicht an die Normstruktur herankommen wie ein individueller Experte, was aber in einem Experiment dieser Art schwerlich zu erreichen sein wird.).

2.2.2. Methode und Durchführung

Aufbau des Experiments. In einer Längsschnittuntersuchung hatten die Versuchspersonen die Aufgabe, 15 Vogelarten anhand vier charakteristischer Merkmalsbereiche zu lernen. In drei Sitzungen, die jeweils im Wochentakt stattfanden, erhielten sie dazu Informationen zu diesen 15 Objekten. Mittels paarweiser Ähnlichkeitsurteile wurde dann die Auswirkung der Lerneinheiten auf die individuellen Wissenskarten der Versuchspersonen überprüft.

Die Versuchspersonen wurden vor dem Experiment zufällig in drei Gruppen aufgeteilt: zwei Experimentalgruppen mit unterschiedlichem Treatment (Exkursionslerner und Kärtchenlerner) und eine Kontrollgruppe³. Die Kärtchenlerner (KL) erhielten in allen drei Lernphasen Kärtchen der 15 Vogelarten mit merkmalsstrukturierten Informationen. Die Exkursionslerner (EL) nahmen im ersten Lerndurchgang mit einem Ornithologieexperten an einer Exkursion teil, an der sie alle 15 Arten beobachten konnten. Der Exkursionsleiter konzentrierte sich in seinen Erläuterungen vornehmlich auf die Merkmalsinformationen, die auch die Kärtchenlerner zur Verfügung hatten. Die beiden folgenden Lernphasen unterschieden sich nicht von denen der Kärtchenlerner. Die Kontrollgruppe durchlief nur die Wissensdiagnose, nahm jedoch nicht an den Lerneinheiten teil. Der Aufbau des Experiments ist in Abbildung 3 skizziert.

³ Ursprünglich sollten mit diesen beiden Lernbedingungen mögliche Unterschiede zwischen einem entdeckenden Lernen (Exkursion) und einem rein instruktionalen Lernen (Kärtchen) aufgedeckt werden. Da sich jedoch, wie die Ergebnisse zeigen werden, die beiden Gruppen nach zufälliger Zuordnung der Versuchspersonen in den Startbedingungen unterschieden, kann auf diese Teilfragestellung in diesem Bericht nicht weiter eingegangen werden.

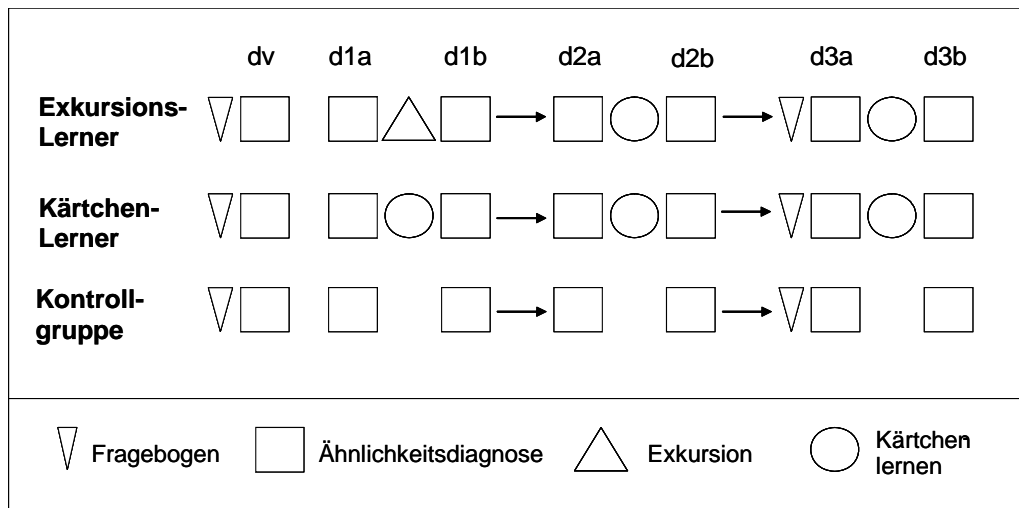


Abbildung 3. Schematische Darstellung des Ablaufs von Experiment 1.

Vor der ersten und vor der dritten Lernphase durchliefen alle Versuchspersonen einen merkmalsorientierten Sachwissenstest, indem ein Multiple-Choice-Fragebogen zu allen 15 Vogelarten die vier Merkmalsbereiche aus den Lerneinheiten testete (Der zweite Test fand mit Absicht nicht unmittelbar nach einer Lerneinheit statt, um Kurzzeiteffekte zu vermeiden. Vor der dritten Lerneinheit lag eine Woche lernfreie Zeit, so dass nur der langfristige Erfolg aus den beiden ersten Durchgängen gemessen wurde.). Vor und nach jeder Lerneinheit wurde der aktuelle Wissensstand aller Versuchspersonen mittels paarweiser Ähnlichkeitsurteile auf der Strukturebene erhoben. Vor der ersten Lerneinheit gaben die Versuchspersonen den Satz von 105 Ähnlichkeitsurteilen zweimal ab, damit sie beim ersten Mal (dv) etwas Übung mit diesem Verfahren bekamen (Der zweite der beiden Datensätze (d1a) wird dann bei der Auswertung als Messung des Vorwissens angesehen). Schließlich wollten wir die Veränderung der Karten eindeutig auf das Lern-Treatment zurückführen und nicht auf einen Gewöhnungseffekt bei den Ähnlichkeitsurteilen. Die Kontrollgruppe durchlief dieselbe Wissensdiagnostik.

Auswahl der Vogelarten und Bestimmung der Zielstruktur. Die Auswahl der 15 Vogelarten, so wie sie sich im Expertenmodell (Abbildung 4) präsentiert, hatte einigen Kriterien zu folgen. (1) Für die Exkursion mussten alle Arten im Raum Zürich im selben Habitat leben und dort leicht zu finden sein, und zwar auch im Winter, als dieses Experiment durchgeführt wurde. (2) Einzelne Arten sollten nicht dermaßen von allen anderen Vögeln verschieden sein, dass auswertungstechnisch betrachtet die Gefahr degenerierter NMDS-Lösungen bestanden hätte. Auch sollten die Arten nicht nur wenigen, scharf getrennten Verwandtschaftsgruppen angehören, um übermäßig kategoriale Lösungen zu vermeiden. Das war technisch wichtig, weil sich sonst ein möglicher Lernfortschritt nicht numerisch fair durch Prokrustes-Transformation von NMDS-Karten messen lassen würde. (3) Ca. sechs Arten sollten Laien (zumindest oberflächlich) bekannt sein, damit eine gewisse Grundstruktur vorhanden war. Mindestens sechs Arten sollten aber den Versuchspersonen völlig unbekannt sein, um ausreichenden Spielraum für das Lernen zu besitzen. Die ersten beiden Restriktionen führten zu einer eingegrenzten Liste möglicher Arten. Diese wurden in einem Vorversuch 30 Studierenden der Universität Zürich in einem Fragebogen vorgelegt, um Bekanntheitsgrad und Vorwis-

sen einschätzen zu können. Daraus erfolgte dann die Auswahl der 15 am geeignetsten erscheinenden Arten.

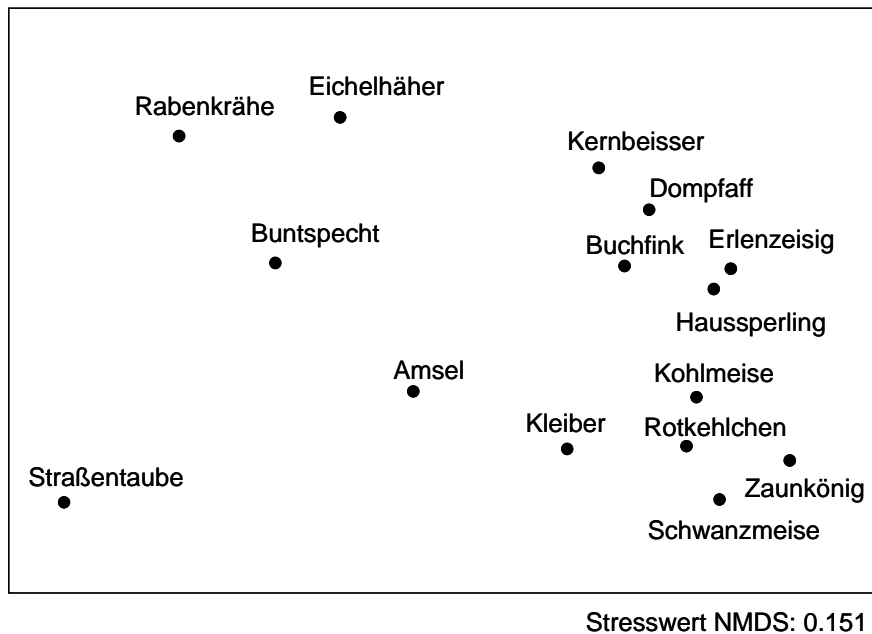


Abbildung 4. NMDS-Karte, die in Experiment 1 die Zielstruktur des Lernens ergibt.

Auf Strukturebene benötigen wir eine Wissensnorm, mit der jede Lernerkarte zu jedem Zeitpunkt verglichen werden kann. Da diese Karte für die numerische Bestimmung des Lernerfolgs sehr wichtig ist, wurde große Mühe darauf verwendet, sie so akkurat wie möglich zu erheben. Deshalb wurden die 15 ausgewählten Arten acht erfahrenen Ornithologie-Experten vorgelegt, die jeweils die paarweisen Ähnlichkeiten einzuschätzen hatten. Bei vier von diesen Experten konvergierten die resultierenden NMDS-Karten in sehr hohem Maße, so dass ihre gemittelten Urteile mit einem Gewicht von 50% zur Grundlage für die Normkarte herangezogen wurden. Um sicherzustellen, dass das Normmodell auch wirklich die zu lernenden Merkmalsbereiche abdeckte, wurden die individuellen Ähnlichkeitsurteile nach dem letzten Lernen für jede der beiden Treatment-Gruppen gemittelt. Jede dieser zwei Ähnlichkeitsmatrizen floss mit einem Gewicht von 25% in die durch NMDS zu bildende Normkarte ein (Abbildung 4).

Lernmaterialien. Die Exkursion fand in zwei Sechsergruppen statt und dauerte ca. eineinhalb Stunden. Die Lerner konnten alle 15 Arten unter Anleitung und begleitenden Kommentaren eines erfahrenen Exkursionsleiters per Fernglas beobachten. Zusätzlich zu seinen Auskünften hatten die Exkursionsteilnehmer die Lernkärtchen in der Hand und konnten sich bei Bedarf daran orientieren. Diese Lernkarten setzten sich je Vogelart aus drei Teilen zusammen: aus a) einer farbigen Abbildung der Vogelart, aus b) einem kurzen, beschreibenden Text, der neben Besonderheiten und leicht eingänglichen Kuriositäten über die Art eine Merkmalsbeschreibung zu Lebensraum, Nahrung und Nistgewohnheiten enthielt (so wie man sie in Standard-Vogelbüchern findet), sowie aus c) einer Merkmalsliste, die alle wichtigen Merkmale aus dem beschreibenden Text tabellarisch aufführte.

In den individuell durchgeführten Lerndurchgängen erhielten die Versuchspersonen jeweils dieselben Lernmaterialien wieder vorgelegt. Sie waren in der Organisation ihres Lernens weitgehend frei, hatten sich aber in den beiden ersten Sitzungen jeweils 30 Minuten und in der dritten Sitzung 20 Minuten damit zu beschäftigen. Als Kriterien wurde ihnen an die Hand gegeben, dass sie versuchen sollten, aufgrund der Abbildungen die Arten zu benennen und dass sie in der verbleibenden Zeit alle Kärtchen mindestens einmal durchgelesen haben sollten. Sie wussten, dass sie Ähnlichkeitsurteile über diese Arten abzugeben hätten und konnten bereits beim Lernen die Kärtchen strukturiert auf einem Tisch ausbreiten.

Versuchspersonen. An der Hauptuntersuchung nahmen 31 Studierende der Universität Zürich im Alter von 22–27 Jahren teil. Sie besaßen keine spezifischen ornithologischen Vorkenntnisse. In den beiden Experimentalgruppen befanden sich jeweils 4 Frauen und 8 Männer, die Kontrollgruppe bestand aus 5 Frauen und 4 Männern. Die Gruppenzuteilung erfolgte vor dem Experiment und zufällig.

2.2.3. Resultate

Merkmalswissen wird gelernt. Vor der ersten und vor der dritten Lerneinheit durchliefen alle Versuchspersonen einen allgemeinen Wissenstest. Zu allen 15 Vogelarten wurde dabei das Merkmalswissen über vier komplexe, mehrere Einzelmerkmale enthaltende Fragen in Form eines Multiple-Choice-Tests erhoben. Maximal konnten in dem Test 60 Punkte erzielt werden. Die Gruppe der Exkursionslerner verbesserte sich von Test 1 zu Test 2 von durchschnittlich 34 auf 46.75 Punkte, die Kärtchenlerner von 30 auf 43.75 Punkte. Beide Verbesserungen sind im *t*-Test auf dem 0.1% -Niveau signifikant (EL: $t(11) = -10.15$, $p < 0.001$, zweiseitig; KL: $t(11) = -6.67$, $p < 0.001$, zweiseitig). Die Versuchspersonen der Kontrollgruppe hingegen verbesserten sich nicht signifikant (von 28 auf 30.77 Punkte; $t(8) = -2.03$, $p = 0.07$, zweiseitig). Im Test 2 sind die beiden Lerngruppen deutlich besser als die Kontrollgruppe (EL: $t(19) = -5.04$; $p < 0.001$; zweiseitig; KL: $t(19) = -3.92$; $p < 0.001$, zweiseitig), unterscheiden sich jedoch nicht signifikant untereinander ($t(22) = 1.28$; $p = 0.21$; zweiseitig). Die erzielten Punktzahlen zeigen deutlich an, dass die beiden Experimentalgruppen in den beiden ersten Lerndurchgängen tatsächlich gelernt hatten, nicht aber die Kontrollpersonen.

Die reine Ratequote in diesem Test lag bei 15 Punkten (60 Fragen mit vier Antwortmöglichkeiten, von denen jeweils eine richtig war). Der Wissenstest dokumentiert also, dass die Versuchspersonen bereits über ein Eingangswissen verfügten. Das war ja auch so geplant, denn es waren bewusst sechs Vogelarten ausgewählt worden, die im Vorversuch bereits bei Laien einen gewissen Bekanntheitsgrad aufwiesen. Gleichzeitig war – wie ebenfalls angezielt – noch nach oben hin genügend Spielraum für eine Verbesserung des Wissens. Die beiden Treatmentgruppen erreichten bis zu Beginn der dritten Lerneinheit etwa die Hälfte des möglichen Wissenszuwachses, so dass in der Längsschnittuntersuchung kein Deckeneffekt auftrat. Mit dieser Überprüfung sind die notwendigen Voraussetzungen hinsichtlich des Merkmalswissens bestätigt: Anders als die Kontrollpersonen verbessern die Lerner ihr Wissen. Damit können nun die Hauptfragestellungen beantwortet werden.

Konvergente Veränderung der Wissenskarten. Gemäß der ersten Hypothese ist zu überprüfen, ob die Wissenskarten der Lerner im Laufe des Prozesses systematisch konvergieren. Wir vergleichen dazu die Wissensstruktur unmittelbar vor dem ersten Lernen (Zeitpunkt d1a) und unmittelbar nach dem dritten Lernen (d3b) miteinander. Dazu werden aus den Ähnlichkeitsurteilen jeder Person zu jedem der beiden ausgewählten Zeitpunkte zweidimensionale NMDS-Karten gerechnet. Diese Karten werden jeweils paarweise mit Prokrustes-Transformation übereinander gelegt. Die resultierende Matrix der AvgLosses lässt sich mit NMDS in einen Raum legen, so dass die Ähnlichkeitsstruktur zwischen den individuellen Wissenskarten sichtbar wird. Letzteres geschieht unter Einschluss der Zielstruktur (um die Richtung einer systematischen Veränderung gleich qualitativ einordnen zu können).

Abbildung 5 zeigt die resultierende NMDS-Karte für die Gruppe der Exkursionslerner. Die Struktur lässt deutlich eine gerichtete Verschiebung erkennen: Die Punkte für die individuellen Wissenskarten zum Zeitpunkt d1b liegen unten links (Ausnahme: EL4-d1a), die zum Zeitpunkt d3b oben rechts (Die Zielstruktur ist in letzterer Hälfte der Lösung enthalten, so dass diese auch die systematisch bessere darstellt.). Neben der Gerichtetheit konvergieren die individuellen Wissenskarten der Exkursionslerner auch mit der Zeit stärker (durchschnittlicher AvgLoss zum Zeitpunkt d1a = 0.70 und zum Zeitpunkt d3b = 0.62; $t(22) = 3.20$, $p < 0.001$, zweiseitig).

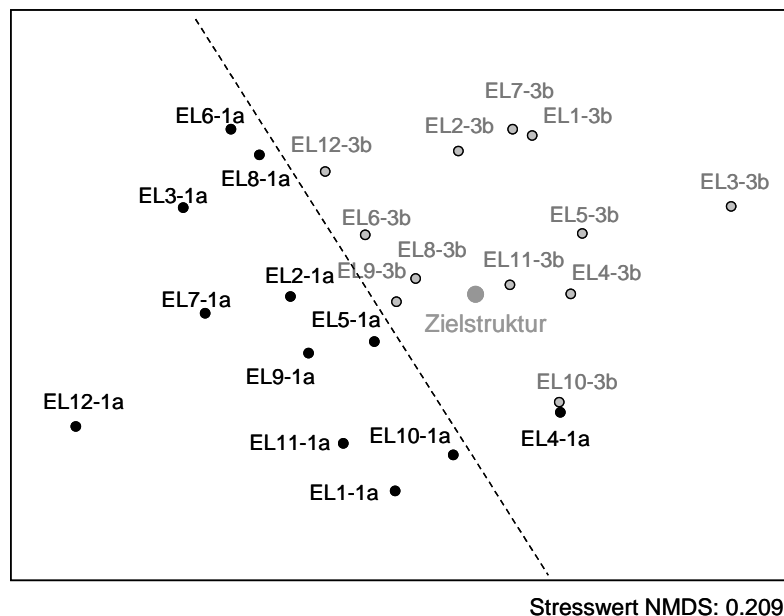


Abbildung 5. NMDS-Lösung der Ähnlichkeiten der individuellen Karten der Exkursionslerner vor dem ersten (d1a) und nach dem dritten Lerndurchgang (d3b).

Die NMDS-Lösung der individuellen Wissensstrukturen der Kärtchenlerner (Abbildung 6) stellt im wesentlichen eine Replikation des oben skizzierten Befundes dar, allerdings nicht in exakt derselben Deutlichkeit: Mit den Kärtchenlernern KL2 und KL8 haben sich bereits vor dem ersten Lernen zwei „heimliche Experten“ eingeschlichen, jedenfalls liegen ihre Wissenskarten verhältnismäßig dicht an der Zielstruktur (ohne dass beide allerdings im Fragebogen herausragend gut abgeschnitten hätten). Auf der anderen Seite bleibt Kärtchenlerner KL4

nach dem dritten Lerndurchgang ein wenig hinter seinen Kollegen zurück. Ohne diese drei Personen könnte man einfach wieder eine senkrechte Trennlinie zwischen d1a und d3b legen.

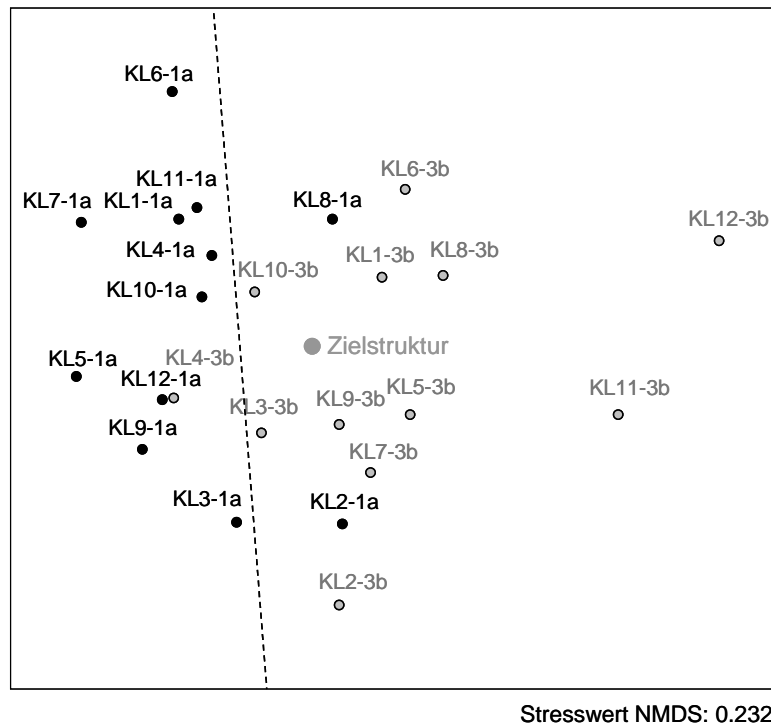


Abbildung 6. NMDS-Lösung der Ähnlichkeiten der individuellen Karten der Kärtchenlerner vor dem ersten (d1a) und nach dem dritten Lerndurchgang (d3b).

Während sich die Gerichtetheit der Veränderungen also repliziert, nimmt die Konvergenz (anders als bei den Exkursionslernern) in der Gruppe der Kärtchenlerner nicht zu (durchschnittlicher AvgLoss zum Zeitpunkt d1a = 0.58 und zum Zeitpunkt d3b = 0.59; $t(22) = -0.51$, $p = 0.60$, zweiseitig). Numerisch liegt das allerdings am Zeitpunkt d1a, zu dem die Karten untereinander deutlich konsistenter sind als in der Gruppe der Exkursionslerner.

Ganz anders sieht die Befundlage bei den Kontrollpersonen aus (Abbildung 7): Hier hat es keine systematische Verschiebung der Wissensstrukturen gegeben. Im Gegenteil, die beiden Karten der jeweils selben Person sind ungefähr im gleichen Quadranten der Lösung wieder zu entdecken. Ohne das Lernen von Merkmalen bleibt die idiosynkratische Strukturierung der Vogelwelt weitgehend bestehen. Qualitativ bewegt sich nur eine einzige Kontrollperson (KP6) systematisch auf die Zielstruktur zu (wobei sich diese Person im Merkmalsfragebogen aber nicht gesteigert hatte). Alle anderen Verschiebungen verlaufen planlos im Raum. Auch eine Zunahme der Konsistenz der Karten untereinander ist nicht zu entdecken (durchschnittlicher AvgLoss zum Zeitpunkt d1a = 0.74 und zum Zeitpunkt d3b = 0.75; $t(16) = -0.44$, $p = 0.65$, zweiseitig).

Damit lässt sich eine systematische Verschiebung bei denjenigen und nur bei denjenigen Personen konstatieren, die die Merkmale der Vogelarten gelernt haben. Im Sinne der aufgestellten Hypothese ergibt sich daraus die Schlussfolgerung, dass das Lernen von Merkmalen die

Bildung von Strukturwissen auf eine systematische Weise beeinflusst. Die erwartete Konvergenz der Karten war jedoch nur bei der Gruppe der Exkursionslerner zu beobachten.

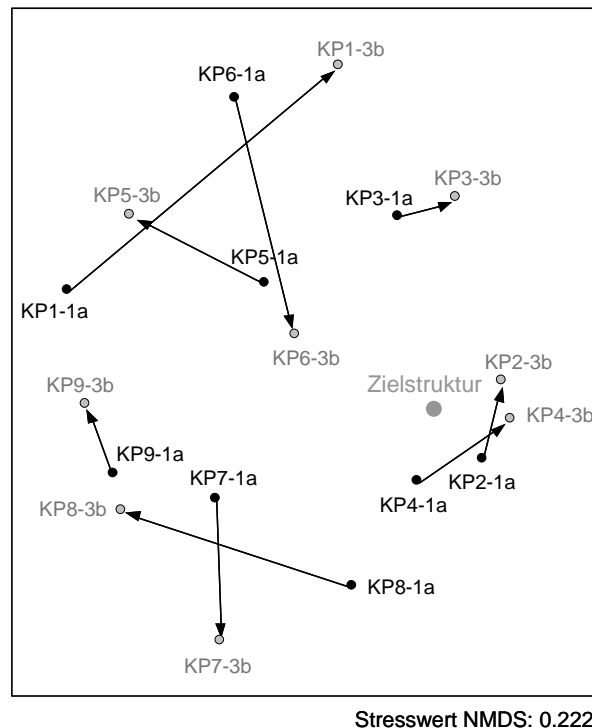


Abbildung 7. NMDS-Lösung der Ähnlichkeiten der individuellen Karten der Kontrollpersonen zum Zeitpunkt d1a und d3b.

Qualitative Verbesserung der Wissenskarten. Aber bedeutet diese systematische Veränderung auch eine systematische Verbesserung? In den vorausgehenden Abbildungen war bereits die Zielstruktur mit eingetragen, um schon einen ersten Hinweis auf die Beantwortung dieser Frage zu haben. Abbildung 8 zeigt dieselbe Auswertelogik für die NMDS-Karten aller gemittelten Ähnlichkeitsmatrizen über alle Zeitpunkte. In dieser Abbildung wird die Zielgerichtetheit besonders deutlich. Immer wenn gelernt wird, bewegt sich die resultierende Karte in Richtung unten links auf das Zielmodell zu.

Wie sich der Wissenszuwachs auf die individuellen Karten der Versuchspersonen auswirkt, ist in Abbildung 9 dargestellt. Die Graphik beinhaltet die strukturelle Differenz zwischen Versuchspersonen und Expertenmodell. Dabei wurde für jede Versuchsperson zu allen Zeitpunkten die Abweichung ihrer individuellen Karte zur Zielstruktur über Prokrustes-Transformation gerechnet (AvgLoss). In der Graphik sind die Mittelwerte der drei Gruppen zu den sechs Messzeitpunkten angegeben.

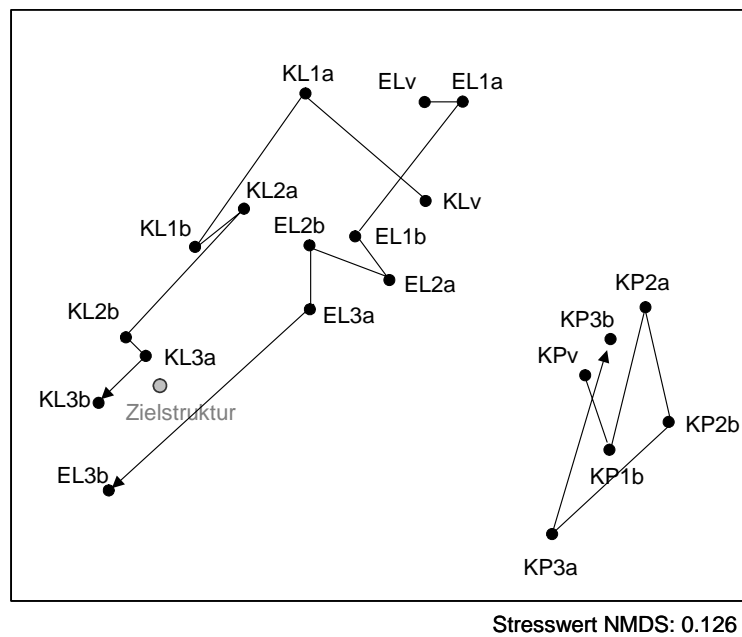


Abbildung 8. Veränderung der Kognitiven Karten (jeweils über die Stichproben gemittelte Proximitäten) über den Lernverlauf (Die Karte d1a bei der Kontrollgruppe degenerierte und ist deswegen nicht skaliert.).

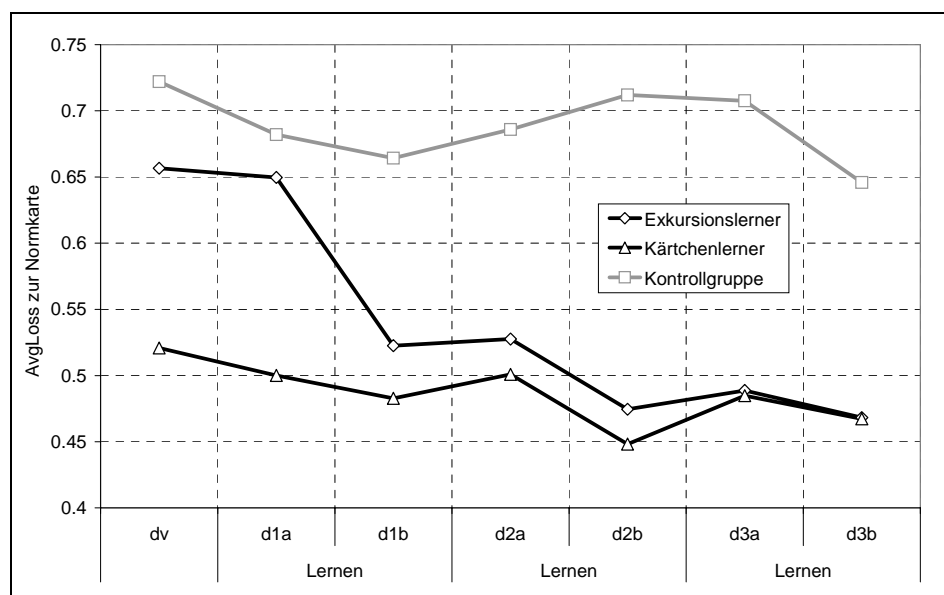


Abbildung 9. Mittlere Differenz (AvgLoss) der individuellen Wissenskarten zur Zielstruktur.

Betrachtet man die Resultate vom Ende her, so besitzt Lernen eine klare positive Auswirkung auf die inhaltliche Qualität der individuellen Wissenskarten: Zum Zeitpunkt d3b, nachdem die beiden Experimentalgruppen drei Lernphasen durchlaufen hatten, sind ihre Strukturen der Expertenstruktur signifikant ähnlicher als die Strukturen der Kontrollpersonen (EL-KP: $t(19) = -3.17, p < 0.01$, zweiseitig; KL-KP: $t(19) = -2.88, p < 0.01$, zweiseitig). Die Karten der nicht lernenden Kontrollpersonen behalten zu allen Messzeitpunkten in etwa dieselbe Abweichung von der Zielstruktur. Zu den Zeitpunkten dv und d1a sind die Strukturen der Exkursionslerner praktisch genauso weit entfernt vom Expertenmodell wie diejenigen der Kontrollpersonen

(Zeitpunkt dv: $t(19) = -1.4, p = 0.17$, zweiseitig; Zeitpunkt d1a: $t(19) = -0.52, p = 0.60$, zweiseitig). Im Unterschied zur Kontrollgruppe gleichen sich ihre Strukturen aber im Verlauf des Lernens nach und nach der Zielstruktur an. Einen grundsätzlich gleichen Lernverlauf kann man auch bei der Gruppe der Kärtchenlerner beobachten. Diese verfügen allerdings zu den Zeitpunkten dv und d1a – aus einem uns nicht erklärbaren Grund – bereits über Wissenskarten, deren AvgLosses signifikant besser sind als die der anderen beiden Gruppen (EL-KL zum Zeitpunkt dv: $t(22) = -4.09, p < 0.001$, zweiseitig und zum Zeitpunkt d1a: $t(22) = -3.14, p < 0.01$, zweiseitig). Unerklärlich ist dieser Befund deswegen, weil er in den Punktzahlen des Multiple-Choice-Tests keinerlei Entsprechung findet. Die genauere Analyse in Abbildung 10 macht deutlich, dass trotz der nur geringen Verbesserung im AvgLoss eine systematische Veränderung der individuellen Karten stattfindet.

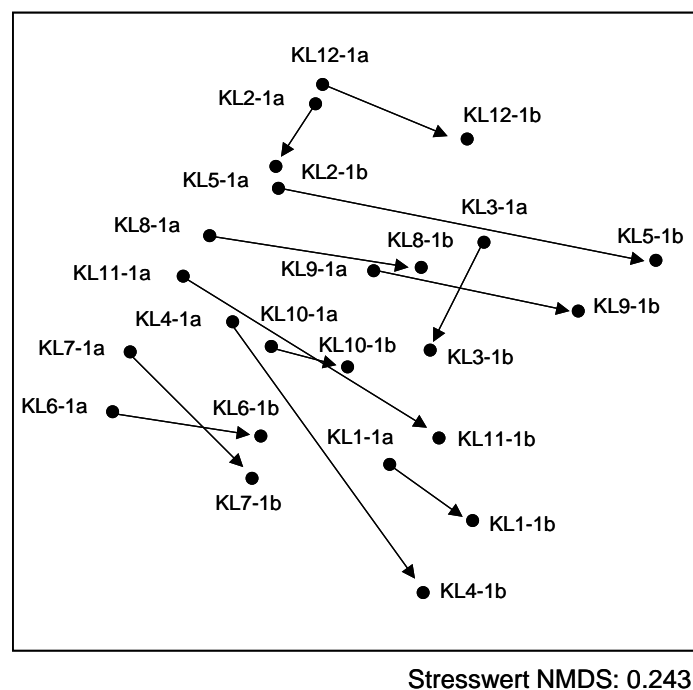


Abbildung 10. NMDS-Lösung der individuellen Karten der Kärtchenlerner zum Zeitpunkt d1a und d1b. Es gibt eine systematische Verschiebung in eine gemeinsame Richtung (nach unten rechts). Alle 12 Karten verschieben sich nach unten, 10 der 12 nach rechts. Obwohl die Karten vom AvgLoss her gesehen nicht nennenswert näher an die Expertenkarte herankommen, ist also trotz allem eine systematische Veränderung zu sehen.

Eine Verbesserung der AvgLoss-Werte fand (wie Abbildung 9 zeigt) jeweils beim Lernen statt, also zwischen der Diagnose d1a und d1b, zwischen d2a und d2b und zwischen d3a und d3b. Während der einwöchigen Pause (d1b zu d2a sowie d2b zu d3a) gab es in beiden Lerngruppen jeweils eine geringe Verschlechterung, die sich als „Vergessen“ interpretieren lässt. Dieser Gesamtbefund (Verbesserung der Karten direkt nach dem Lernen, Verschlechterung nach einer Woche Pause) ist mit einem Binomialtest auf dem 1%-Niveau gegen die Zufallserwartung von .50 sowohl für die Exkursionslerner als auch für die Kärtchenlerner abgesichert (jeweils $N = 60$, fünf prognostizierte „Abs“ und „Aufs“ für jeweils 12 Versuchspersonen gingen in die Verlaufsüberprüfung jeder Lerngruppe ein). Für die Kontrollpersonen ergibt sich kein signifikantes Ergebnis.

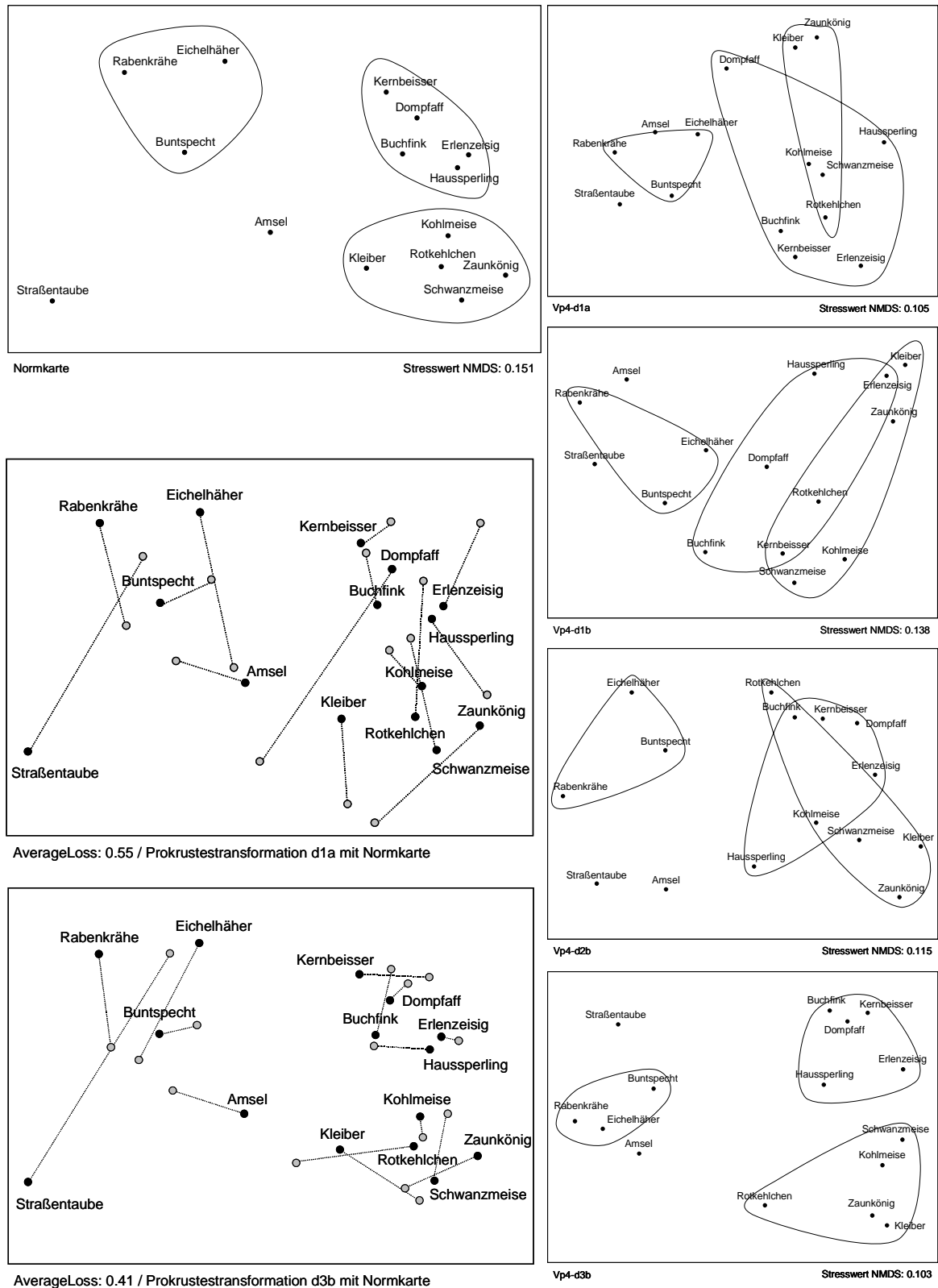


Abbildung 11. Lernfortschritt eines individuellen Lerners (d1a, d1b, d2b, d3b) mit eingezeichneten Expertenclustern und mit Prokrustes-Transformation der Karten d1a und d3b mit der Normkarte.

Wirklich dramatisch ist als Einzeleffekt allerdings nur die Verbesserung der Kognitiven Karten der Exkursionslerner im ersten Lerndurchgang, die beim Mittelwertvergleich im t -Test signifikant wird ($t(22) = 4.06, p < 0.001$, einseitig). Offensichtlich geschieht das wesentliche

Lernen bereits im ersten Durchgang, und da die Kärtchenlerner bereits vorher recht gut waren (aus, wie gesagt, für uns nicht nachvollziehbaren Gründen), fällt bei ihnen dieser große Strukturverbesserungseffekt aus. Abbildung 11 zeigt eine die Karten einer durchschnittlichen Exkursionslernerin vor und nach dem Lernen. Das moderate Absinken des AvgLosses von 0.55 auf 0.41 geht mit einer sichtbaren Strukturverbesserung einher: Die zusammengehörenden Gruppen der Vogelarten finden zusammen, und die Abweichungen zum Normmodell wirken sich nach dem Lernen nur noch innerhalb der Cluster aus (Die alternative Auswertung, welche die Cluster aus der Normkarte in die Lernerkarte legt, zeigt sehr schön, wie sich die relevanten Gruppen finden.). Die größte Abweichung besteht nach dem Lernen bei der Straßentaube, was aber keinen sonderlich hohen Stellenwert besitzt, da sie sowohl in der Normkarte als auch in der Lernerkarte eine Einzelposition bezieht.

2.2.4. Diskussion

Der Hauptbefund des Experiments entspricht den Erwartungen an die individuellen Wissenskarten: Sie verbessern sich mit dem Lernen systematisch auf die Zielstruktur zu, während in der einwöchigen Pause zwischen den Lernsitzen jeweils wieder etwas von der Qualität der Karten verloren geht. Dass die individuellen Karten das Zielmodell nach drei Lerndurchgängen nicht vollständig erreichen (und im Durchschnitt weiter von diesem weg bleiben als die befragten langjährigen Experten), lässt sich mit der Schwierigkeit der Materie erklären: Die Punktzahlen in den Multiple-Choice-Fragebögen signalisieren noch ausreichend Spielraum nach oben hin, so dass die verbleibende Diskrepanz zur Zielstruktur zumindest zum Teil auf ein noch mangelhaftes Wissen zurückgeführt werden kann. Dass jedoch auch die Experten individuell deutlich vom Zielmodell abweichen, spricht vielleicht für die Komplexität des gewählten Wissensgebietes oder aber dafür, dass dieses Wissensgebiet eben auch Raum für idiosynkratische Strukturvorstellungen lässt. Da vergleichende Untersuchungen bislang völlig fehlen, ist als Alternativerklärung aber auch nicht auszuschließen, dass das technische Verfahren (Ähnlichkeitsurteile, NMDS und Prokrustes-Transformation) gar keine wirklich gut aufeinander passenden individuellen Karten zu produzieren in der Lage ist. Deswegen sollte Experiment 2 versuchen, die ersten Erklärungen (Komplexität und unvollständige Konvergenz des Wissensgebietes) nach Möglichkeit auszuschalten, um die Grenze des durch das Verfahren technisch Machbaren etwas genauer auszuloten.

Dass sich die individuellen Wissenskarten zwar systematisch verbessern, dabei jedoch nur in der Gruppe der Exkursionslerner untereinander konvergieren, hängt mit dem unerklärlich guten Ausgangswert der Kärtchenlerner zusammen. Aber auch für die Exkursionslerner ist der Effekt trotz statistischer Signifikanz numerisch geringer als erwartet (von einem AvgLoss von 0.70 hin zu einem AvgLoss von 0.62). Das könnte damit zusammenhängen, dass die Versuchspersonen in der Regel sechs der 15 Arten bereits kannten und hinsichtlich der verbleibenden neun Arten zwar inadäquate, aber doch teilweise konvergierende Vorstellungen hatten. Auf diese Weise könnte es zur beobachteten „Parallelverschiebung“ der individuellen Karten durch das Lernen gekommen sein. Natürlich ist das eine in den Details schwer zu überprüfende Interpretation, doch eine letztlich Klärung können die vorliegenden Daten wohl

nicht bringen. Wir setzen damit unsere Hoffnungen hinsichtlich der noch zu zeigenden Konvergenz auf Experiment 2.

Der größte Sprung in der Verbesserung der Wissenskarten wird durch die erste Lerneinheit ausgelöst, besonders ausgeprägt zu sehen für die Gruppe der Exkursionslerner. Dieser Effekt kann zweierlei bedeuten: Entweder ist das „Aha“-Erlebnis, wie die einzelnen Vögel zueinander stehen, bereits durch das Kennen lernen der 15 Arten hinlänglich gegeben. Oder aber die Merkmalsinformationen werden im zweiten und dritten Lerndurchgang lediglich wieder aufgefrischt, es wird aber eigentlich nichts mehr dazugelernt. Beide Erklärungen widersprechen sich nicht unbedingt und sind aus den vorliegenden Daten nicht überprüfbar. Es bleibt also eine offene Frage für zukünftige Forschung, wie viel Merkmalsinformation über die einzelnen Objekte bekannt sein muss, damit eine Person mit ihren Ähnlichkeitsurteilen eine in den Grundzügen stimmige Wissenskarte produzieren kann.

2.3. Experiment 2

2.3.1. Fragestellung und Hypothesen

Experiment 1 hat gezeigt, dass sich die Wissenskarten merkmalslernender Personen verändern und verbessern. Allerdings konnte dabei keine signifikante Konvergenz der Karten festgestellt werden, und wegen des doch recht anspruchsvollen Lernmaterials fand auch keine perfekte Angleichung der Karten an den Stand der Experten statt. Das zweite Experiment soll deswegen – parallel aufgebaut, jedoch mit einfacher gestricktem und perfekt konvergentem Lernmaterial – ausloten, wie stark die mögliche Annäherung an eine „perfekte“ Wissenskarte sein kann.

2.3.2. Methode und Durchführung

Aufbau des Experiments. In einer Längsschnittuntersuchung hatten die Versuchspersonen die Aufgabe, zu 15 Ländern die drei Merkmale Landesfläche, Einwohnerzahl und Bruttosozialprodukt so zu lernen, dass sie daraus eine Wissenskarte aufbauen konnten. In vier Sitzungen, die jeweils im Wochentakt stattfanden, erhielten sie dazu Informationen zu diesen 15 Objekten. Mittels paarweiser Ähnlichkeitsurteile unmittelbar vor und unmittelbar nach der Lerneinheit wurde dann die Auswirkung der Lerneinheiten auf die individuellen Wissenskarten der Versuchspersonen überprüft. Zu Beginn jeder Sitzung erhob ein computerbasierter Multiple-Choice-Fragebogen den aktuellen Wissensstand über die drei Merkmale zu jedem Land. In diesem Fragebogen hatten die Versuchspersonen die Länder einer Liste aller einzelnen Merkmalsausprägungen (plus Distraktoren) zuzuordnen. Zu jeder Angabe war zudem ein Konfidenzurteil abzugeben, so dass der Fragebogen zwischen lückenhaftem Wissen und falschem Wissen (die Angabe ist inkorrekt, die Person ist sich aber der Korrektheit subjektiv sicher) unterscheiden konnte.

Die Versuchspersonen wurden vor dem Experiment zufällig in eine Experimentalgruppe und eine Kontrollgruppe aufgeteilt. Die Versuchspersonen der Kontrollgruppe lernten nicht, son-

dern führten lediglich das komplette Programm der Wissensdiagnostik durch. Der Aufbau des Experiments ist in Abbildung 12 skizziert.

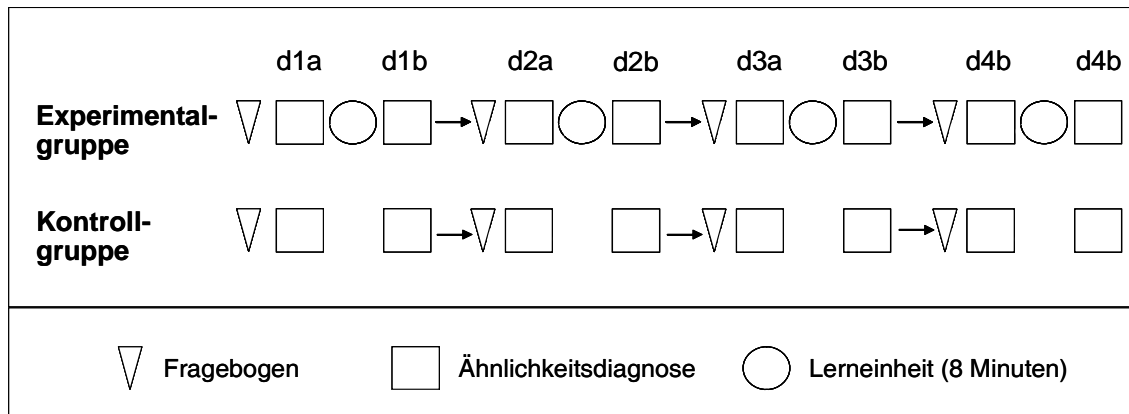


Abbildung 12. Schematische Darstellung des Ablaufs von Experiment 2.

Auswahl der Länder und Bestimmung der Zielstruktur. In einem Vorversuch mit insgesamt 46 Versuchspersonen hatte sich aus einer Vorauswahl von 26 Ländern der Pool von 15 Ländern als gut geeignet herausgestellt, der zur Grundlage der Zielstruktur (Abbildung 13) gemacht wurde. Außerdem hatten sich die drei intervallskalierten Merkmale Landesfläche, Einwohnerzahl und Bruttosozialprodukt pro Kopf als günstig für Lernen, Wissensdiagnose und Zielstruktur erwiesen. Die Zielstruktur errechnete sich über die Unähnlichkeiten der einzelnen Länder auf diesen drei Merkmalen. Dazu wurde folgende Formel verwendet:

$$U\ddot{A}_{(x,y)} = \frac{\left| (x_i - y_i) / (x_i + y_i) \right| + \left| (x_j - y_j) / (x_j + y_j) \right| + \left| (x_k - y_k) / (x_k + y_k) \right|}{3}$$

Wobei: $U\ddot{A}_{(x,y)}$: Unähnlichkeitsquotient eines Länderpaares
 x, y : Länderindex
 i : Einwohnerzahl
 j : Fläche in km²
 k : BSP pro Kopf in US\$

Dieser Quotient $U\ddot{A}_{(x,y)}$ beschreibt den Betrag des prozentualen Unterschiedes je zweier Länder aufgrund ihrer Merkmalsausprägungen. Die resultierende Dreiecksmatrix wurde mittels NMDS (RobuScal-Algorithmus) in eine zweidimensionale Karte transformiert, die für die Qualitätsabschätzung der individuellen Wissenskarten jeweils als Norm gilt (Abbildung 13).

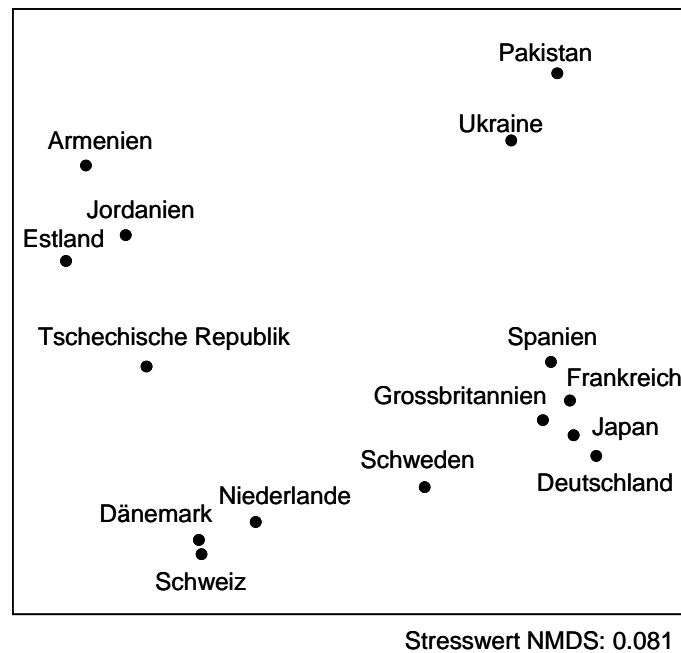


Abbildung 13. NMDS-Karte, die in Experiment 2 die Zielstruktur des Lernens ergibt.

Lernmaterialien. Das Lernmaterial bestand aus 15 Kärtchen, auf denen zu je einem Land die Angaben über Einwohnerzahl, Fläche und Bruttosozialprodukt standen. Die Versuchspersonen waren frei in ihrer Lernstrategie; sie konnten die Lernkarten ordnen und Gruppen bilden, ausbreiten usw. Die Lerneinheit dauerte jeweils acht Minuten.

Versuchspersonen. 37 Personen (17 weiblich; 20 männlich) im Alter zwischen 19 und 33 Jahren nahmen an dem Experiment teil. Es handelte sich dabei hauptsächlich um Studentinnen und Studenten der Psychologie, Soziologie, Tiermedizin und Pharmazie. Hinzu kamen einige Versuchspersonen ohne Hochschulausbildung. 25 Personen wurden der Experimentalgruppe zugeordnet, 12 der Kontrollgruppe. Die Variablen Alter, Geschlecht und Schulbildung waren in beiden Gruppen gleichverteilt.

2.3.3. Resultate

Merkmalswissen wird gelernt. Zu Beginn jeder der vier Sitzungen wurde das Merkmalswissen der Versuchspersonen mit einem Zuordnungs-Fragebogen geprüft. Dabei galt ein Datum als gewusst, wenn die Angabe innerhalb einer 20%-Marke des tatsächlichen Wertes gemacht wurde. Abbildung 14 zeigt, dass für jedes der drei Merkmale in der Experimentalgruppe ein deutlicher Wissenszuwachs zu verzeichnen ist. Dabei basiert dieser Fortschritt vor allem auf einem „Auffüllen“ von Wissenslücken: Die Zunahme des korrekten Wissens und die gleichzeitige Abnahme der Wissenslücken zeigen für jedes der drei Merkmale in einer zweifaktoriellen Varianzanalyse (ANOVA) mit den zwei Intersubjektfaktoren Merkmal und Zeit signifikante Effekte des Merkmals, $F(1.78, 42.82) = 10.14$, $MSE = 86.30$, $p < .001$, und der Zeit, $F(2.0, 47.98) = 53.57$, $MSE = 457.44$, $p < .001$. Dabei wird von vornherein das Bruttosozialprodukt etwas besser gewusst als die Einwohnerzahl, und diese wiederum besser als die Fläche des jeweiligen Landes. Der Anteil falschen Wissens bleibt über alle Durchgänge und über alle Merkmale hinweg erstaunlich konstant bei ca. 20% des Gesamtwissens.

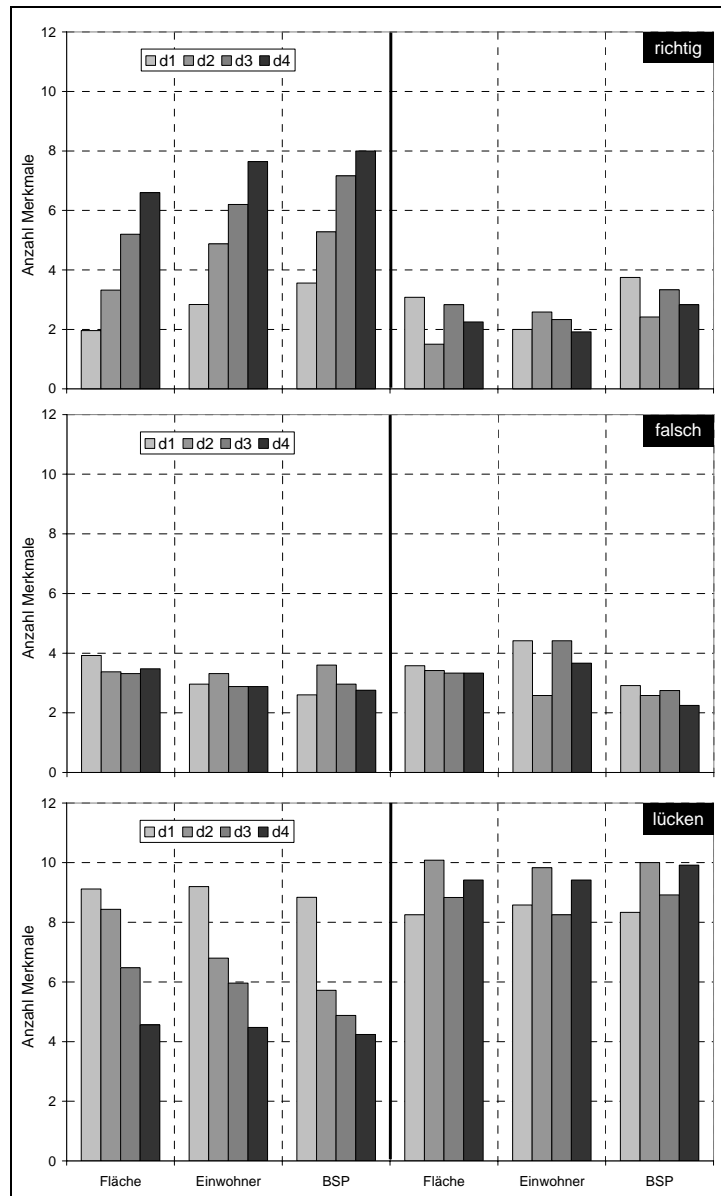


Abbildung 14. Verlauf des Merkmalswissens der Experimentalgruppe (links) und der Kontrollgruppe (rechts) in Experiment 2. Angegeben sind die durchschnittlichen Anzahlen pro Versuchsperson über alle 15 Länder; d1 – d4 = Erhebungszeitpunkte.

Bei der Kontrollgruppe ergibt sich keinerlei signifikante Veränderung. Diese Versuchspersonen steigen auf demselben Wissensniveau ein wie die Personen der Lerngruppe, verbleiben dort jedoch auch. Die Lerner sind bereits ab dem zweiten Durchgang signifikant besser, und diese Schere geht von Durchgang zu Durchgang weiter auf (d1: $t(109) = -0.33, p = 0.37$; d2: $t(91) = 4.79, p < 0.001$; d3: $t(99) = 6.23, p < 0.001$; d4: $t(106) = 9.14, p < 0.001$; einseitig).

Insgesamt zeigt Abbildung 14, dass die Versuchspersonen ihre Ähnlichkeitsurteile durchaus nicht auf dem Niveau perfekten Wissens abgeben (zumindest bis zum Zeitpunkt d4a, denn nach dem vierten Lerndurchgang wurden ja keine Fragebogendaten mehr erhoben). Die im folgenden zu betrachtenden Wissenskarten müssen also vor dem Hintergrund gelesen werden, dass den Versuchspersonen – selbst wenn sie ihr Wissen durch die Lerndurchgänge fast ver-

dreifacht haben – noch immer ca. 30% der Informationen unbekannt sind und dass sie bei 20% des Gesamtwissens von inadäquaten Vorstellungen ausgehen.

Konvergente Veränderung der Wissenskarten. Wie in Experiment 1, so wurden auch hier alle individuellen Ähnlichkeitsmatrizen in zweidimensionale NMDS-Karten überführt und paarweise einer Prokrustes-Transformation unterzogen. Die Dreiecksmatrix der resultierenden AvgLoss-Werte zwischen allen individuellen Karten der Lernergruppe aus Durchgang d1a (vor dem ersten Lernen) und Durchgang d4b (nach dem vierten Lernen) wurde mit NMDS als zweidimensionaler Raum skaliert. Abbildung 15 zeigt die systematische Verschiebung der Wissensstrukturen nach dem Lernen in die rechte Region der Karte. Lediglich zwei (dunkelgrau markierte) Wissenskarten bleiben beim Niveau vor dem Lernen, und die Wissenskarten von vier Versuchspersonen (plus eine Kontrollperson) hatten sich bereits von Anfang an in der Zielregion befunden (zur Verdeutlichung dieses Befundes ist eine „Demarkationslinie“ in die Abbildung eingetragen). Die Kontrollpersonen positionieren sich in der Region der Experimentalgruppe und zeigen keine systematische Verschiebung.

Anders als in Experiment 1 zeigt sich neben dieser systematischen Verschiebung auch eine Konvergenz der individuellen Wissensstrukturen: Die mittlere Distanz zwischen allen Wissenskarten beträgt vor dem Lernen 0.50, nach dem Lernen 0.38. Dieser Unterschied ist im t -Test signifikant ($t(299) = 15.12, p < 0.001$). Für die Wissensstrukturen der Kontrollpersonen findet sich weder eine systematische Verschiebung noch eine Konvergenz. Damit bestätigt sich für das herangezogene Wissensgebiet der Länder, dass die individuellen Karten durch das Merkmalslernen systematisch konvergieren.

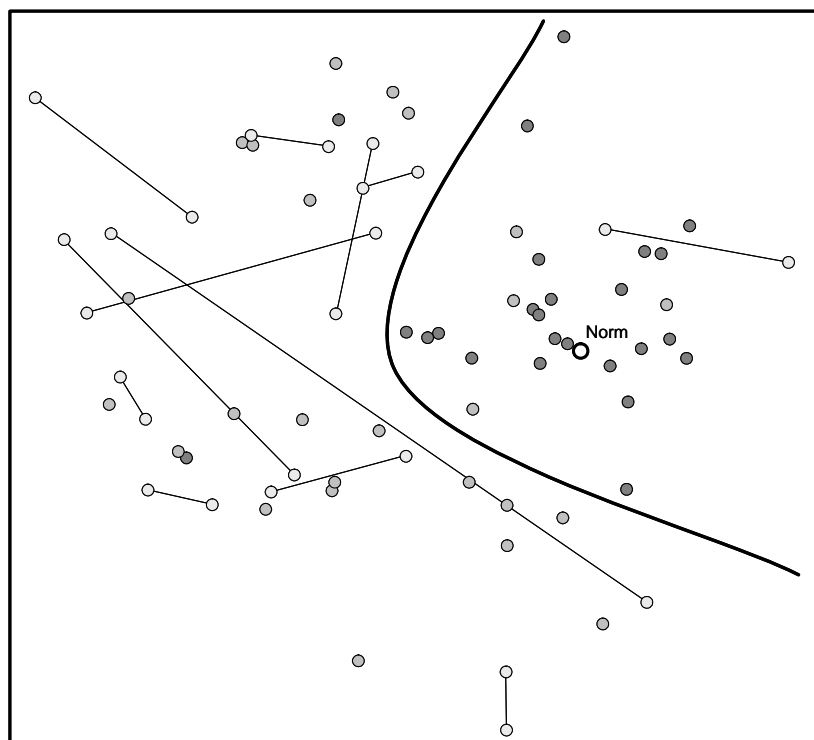


Abbildung 15. NMDS-Lösung mit allen individuellen Wissenskarten vor dem ersten Lernen (d1a = grau) und nach dem vierten Lernen (d4b = dunkelgrau). Zusätzlich ist zur Orientierung die Zielstruktur des Lernens (Norm) eingetragen. Die verbundenen, hellgrauen Punkte stehen für die Karten der Kontrollgruppe zum Zeitpunkt d1a und d4b.

Qualitative Verbesserung der Wissenskarten. In Abbildung 15 ist bereits die Zielstruktur des Lernens mit eingetragen. Aus der systematischen Verschiebung auf diese Zielstruktur hin ergibt sich bereits ein erster Hinweis auf die qualitative Verbesserung der Wissenskarten nach dem Lernen. Dieser wird durch eine Analyse aller AvgLosses zur Zielstruktur untermauert (Abbildung 16): Nach jeder der vier Lerneinheiten nähern sich die individuellen Wissenskarten unmittelbar der Zielstruktur an, so dass sie am Ende des Experiments bei einem AvgLoss von durchschnittlich 0.29 enden (der Unterschied zum Start des Experiments ist im t -Test auf dem 0,1%-Niveau signifikant; $t(24) = 6.26, p < 0.001$). Interessant ist auch, dass ein Teil der Verbesserung in den Wissenskarten in der Woche zwischen den Sitzungen jeweils wieder verloren geht. In Sitzung 3 war der Lerneffekt so gering, dass davon zu Beginn der vierten Sitzung nichts mehr übrig ist. Ansonsten überwiegt aber der Lerneffekt deutlich den Vergessenseffekt. Das Muster von Verbesserung der Wissenskarten durch Merkmalslernen und Verschlechterung durch eine Woche Pause wird im Binomialtest auf dem 0,1%-Niveau signifikant.

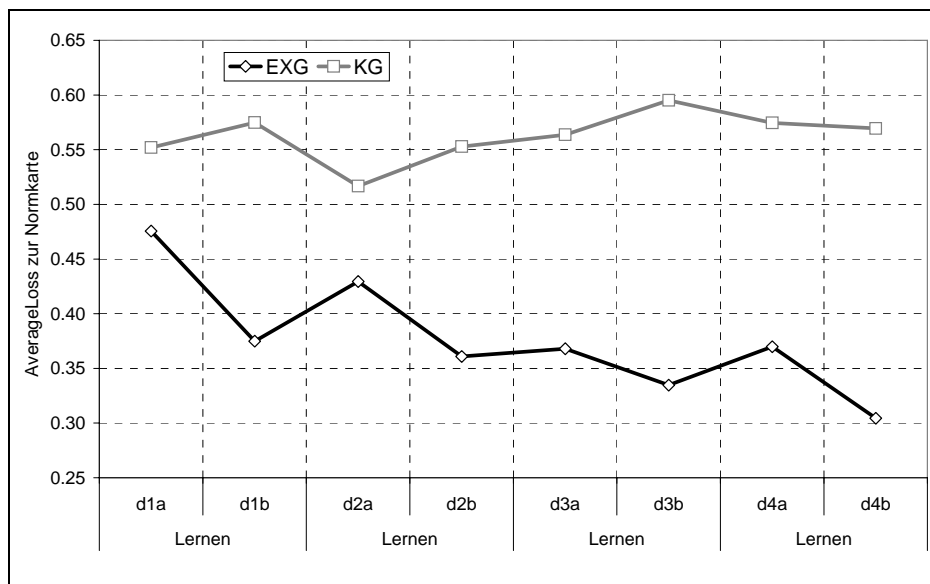


Abbildung 16. Gemittelte Abweichungen aller individuellen Wissenskarten von der Zielstruktur.

Die Versuchspersonen der Kontrollgruppe hingegen bilden durchgängig schlechtere Karten als die Lernergruppe (einzig die Unterschiede beim Ausgangswert in d1a sind nicht signifikant; $t(35) = -1.53, p = 0.07$, einseitig). Die gemittelten AvgLosses schwanken unsystematisch um den Ausgangswert von 0.55, ohne sich systematisch zu verbessern oder zu verschlechtern (t -Test zwischen d1a und d4b: $t(11) = -0.38, p = 0.35$). Das mehrfache Abgeben von Ähnlichkeitsurteilen allein ergibt also keine Verbesserung in den Wissenskarten.

Der Lernfortschritt einer überdurchschnittlichen Person aus der Lernergruppe ist in Abbildung 17 illustriert (aufgebaut analog zu Abbildung 10). Die große Verbesserung im AvgLoss (von 0.59 auf nur noch 0.22) korrespondiert mit einer Karte, in der nur noch Japan maßgeblich von der Zielposition entfernt liegt, indem der Lerner dieses asiatische Land nicht vollständig in die Gruppe der ansonsten europäischen großen Industrienationen integriert. Das Einzeichnen der Cluster aus der Normkarte macht deutlich, dass der Lerner die grundsätzliche Zweiteilung

bereits nach dem ersten Lernen (AvgLoss 0.39) und die weitere Unterteilung in vier Cluster bereits nach dem zweiten Lernen (AvgLoss 0.28) in seiner Wissenskarte repräsentiert hatte. Die qualitativen Beobachtungen bei der Interpretation der Karte korrespondieren damit sehr gut mit dem quantitativen Maß des AvgLoss und verdeutlichen die zielgerichtete Annäherung des strukturellen Wissens des Lernenden hin zur normativen Zielstruktur.

2.3.4. Diskussion

Auch Experiment 2 bestätigt die Hypothese, dass sich Wissenskarten durch das Lernen einzelner Merkmale substantiell verbessern. Zusätzlich ist es durch die Auswahl des Wissensgebietes und des Lernmaterials gelungen, den gewünschten Konvergenz-Effekt zwischen den individuellen Karten zu erzielen. Die Wissenskarten liegen außerdem – wie aufgrund des vereinfachten Materials angenommen – wesentlich dichter an der Zielstruktur als bei dem sehr komplexen Wissensfeld aus Experiment 1.

Gleichwohl bleibt eine mittlere AvgLoss-Distanz von 0.30 zur Zielstruktur bestehen. Ein Teil dieses Unterschiedes lässt sich sicherlich darauf zurückführen, dass die Versuchspersonen selbst vor dem vierten Lerndurchgang (d.h. im letzten der vier Fragebögen) nur über ein korrektes Merkmalswissen von ca. 50% verfügten. Zukünftige Forschung könnte hier also mit einem strengerem Kriteriumslernen noch weiter ausloten, wie nahe man Lerner durch reines Merkmalslernen an eine integrierte Zielstruktur heranführen kann. Als unsystematischer Hinweis dazu mag gelten, dass wir drei Personen einmal die Kärtchen mit den Merkmalen gegeben haben und sie baten, aufgrund dieser vorliegenden Informationen paarweise Ähnlichkeitsurteile zwischen den 15 Ländern zu bilden. Die entstandenen drei „Wissens“-karten wichen durchschnittlich mit einem AvgLoss von 0.16 von der Zielstruktur ab. Aber hier handelt es sich um gegebene Information, und nicht um integriertes Wissen. Aus diesem Grund wären wir skeptisch, ob derart von Versuchspersonen konstruierte Ähnlichkeitsurteile auf dieselbe Weise zustande kämen. Trotzdem sollten die gemessenen Werte ermuntern, nach einer weiteren Annäherung an eine Zielstruktur zu suchen.

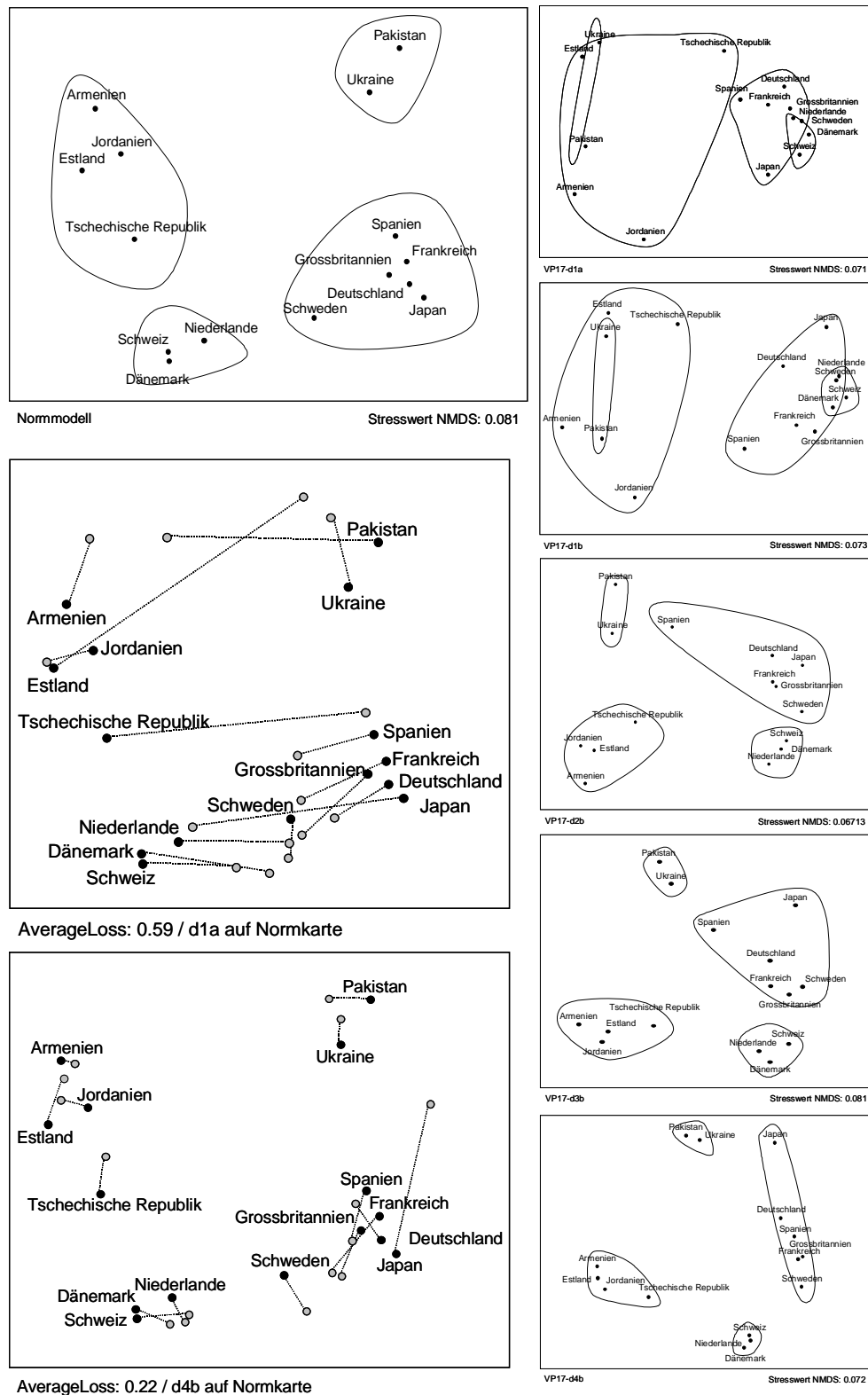


Abbildung 17. Lernfortschritt eines individuellen Lerners (1a,1b,2b,3b,4b) mit eingezeichneten Experten-Cluster und mit Prokrustes-Transformation der Karten d1a und d4b mit der Normkarte.

2.4. Gesamtdiskussion

Die zwei präsentierten Experimente haben aufgezeigt, dass sich das Lernen von Merkmalsinformationen auf die Ähnlichkeitseinschätzung von Objekten auswirkt und dieses wiederum einen systematischen strukturierenden Einfluss auf die Wissenskarten einer Person hat. Damit bietet sich das vorgeschlagene Verfahren – die Umsetzung einer Matrix von paarweisen Ähnlichkeitsurteilen in eine NMDS-Lösung und nachfolgende Prokrustes-Transformation dieser Karte mit einer Zielstruktur oder einem Expertenmodell – als generelles Werkzeug für die Messung des Wissensgrades einer Person in objektbasierten Inhaltsgebieten an. Hier kann nicht nur der AvgLoss als eine quantitative Größe für die qualitative Abweichung des Wissens einer Person von einer gesetzten Norm verwendet werden. Die dargestellten Einzelfallbeispiele untermauern auch die intuitive Interpretierbarkeit der strukturellen Details solcher NMDS-Lösungen.

Unseres Wissens handelt es sich bei den vorgelegten Experimenten um die erste systematische Studie, die den Einfluss von Lernen auf NMDS-Wissenskarten untersucht und dabei unterschiedliche Wissensstände zu unterschiedlichen Zeitpunkten miteinander vergleicht. Der bisherige Standard in der psychologischen und sozialwissenschaftlichen Forschung war der Einsatz dieses Skalierungsverfahrens für gemittelte Stichprobendaten. Das hatte, wie bereits in der Einleitung ausgeführt, seinen guten technischen Grund in der mangelhaften Robustheit (und damit mangelhaften Reliabilität) der vorliegenden Algorithmen. Durch den RobuScal-Algorithmus wird aber nun die Skalierung individueller Datensätze auch dann möglich, wenn diese möglicherweise mit Ausreißern behaftet sind. Das eröffnet dem Verfahren Perspektiven für all jene Wissensbereiche, die sich durch zueinander in Relation stehende Objekte definieren.

Dabei verstehen wir die Wissenskarten als Ausdruck einer Messung, und nicht als eine explizite Vorstellung, über die eine Person in dieser Form verfügt (Dass Letzterem so ist, drückt sich in der Schwierigkeit aus, die eine Person hat, wenn man sie bittet, zu einem bestimmten Gebiet ihre Wissenskarte zu zeichnen – in der Regel gelingt dies entweder gar nicht oder nur sehr fehlerhaft.). Komplexe Wissenskarten können deswegen mit dem Verfahren von Ähnlichkeitsurteilen „gezeichnet“ werden, weil diese Paarurteile dasjenige sequenzialisiert abfragen, was im Resultat dann parallel vorliegt. Derartig komplexes relationales Wissen bleibt für die Person selbst in der Regel etwas Implizites. Lediglich einfaches kategoriales Wissen (wie die Zuordnung von Tier- oder Pflanzenarten zu zoologischen oder botanischen Gruppen) oder klar dimensional strukturiertes relationales Wissen (wie z.B. die Links-Rechts-Achse für die Verortung politischer Parteien) ist in seiner Relationalität einer Person mit entsprechender Expertise auch bewusst. Nachdem die vorliegenden Experimente den qualitativen Zusammenhang dieser Karten mit dem Merkmalswissen gezeigt haben, ist die Möglichkeit des Explizitmachens strukturell sonst zumeist nicht bewusster Anteile des integrierten Wissens ein weiteres Argument, das vorgeschlagene Verfahren in der Wissensdiagnostik einzusetzen.

Ein drittes Argument schließlich ist die Form des Wissens, das gemessen wird. Durch das Abgeben von Ähnlichkeitsurteilen wird von den getesteten Personen eine Integrationsleistung ihres Merkmalswissens erwartet, die nicht erbracht werden kann, wenn Wissen lediglich stu-

pide auswendig gelernt wurde. Insofern hat die Messung von Wissenskarten auch gegenüber gleich gut automatisierbaren Multiple-Choice-Tests einen Vorteil: Die paarweise Einschätzung von Ähnlichkeiten zwischen zwei Objekten stellt bereits die erste „Anwendung“ des erlernten Merkmalswissens dar, und diese Anwendung kann nur in der integrierten Form eines Gesamturteils korrekt erfolgen, wenn die einzelnen Daten auch verstanden wurden.

Diese Argumente sprechen für einen Einsatz von NMDS-Wissenskarten in der diagnostischen Praxis. Auch wenn die vorgelegten Experimente die grundsätzliche Abbildungsleistung von Merkmalswissen belegen konnten, bleibt dennoch weitere Forschung notwendig. Zu allererst sollte ausgetestet werden, wie weit sich die Interpretation der Kartengüte nicht nur auf die Gesamtqualität beziehen lässt, sondern auch auf Details in der Karte ausgeweitet werden kann. Ein Schritt in diese Richtung könnte der Abbildungsfehler der einzelnen Objekte in der Prokrustes-Transformation sein (vgl. Kapitel 4 in dieser Arbeit). Des Weiteren sollte mit noch stärker kontrollierten Lernexperimenten gezeigt werden, wie dicht derartige individuelle Karten an eine Wissensnorm herangeführt werden können. Außerdem verweist Läge (2004) darauf, dass es auch nicht-konvergente Wissensgebiete gibt, für die es nicht möglich sein dürfte, eine Wissensnorm in Form einer eindeutigen Karte zu formulieren. Hier müssten andere Mittel entwickelt werden, um die Güte der Karte zu testen. Dass sich individuelle Wissenskarten aber systematisch mit dem Stand von Merkmalswissen verändern, ist ein erster wichtiger Schritt auf dem Weg zu einem neuen Messverfahren in der Wissenspsychologie.

2.5. Literatur

- Aebli, H. (1976). *Psychologische Didaktik. Didaktische Auswertungen der Psychologie von Jean Piaget*. Stuttgart: Klett.
- Ausubel, D. P. (2000). *The acquisition and retention of knowledge*. Dordrecht: Kluwer Academic Publishers.
- Bonato, M. (1990). *Wissensstrukturierung mittels Struktur-Lege-Techniken. Eine graphen-theoretische Analyse von Wissensnetzen*. Frankfurt a.M.: Lang.
- Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling – Theory and applications*. New York: Springer.
- Eckert, A. (1998). *Kognition und Wissensdiagnose. Die Entwicklung und empirische Überprüfung des computergestützten wissensdiagnostischen Instrumentariums „Netzwerk-Elaborierungs-Technik (NET)“*. Lengerich: Pabst.
- Eckert, A. (1999). Die „Mannheimer Netzwerk-Elaborierungs-Technik (MaNET)“ – Ein computergestütztes Instrument zur Analyse vernetzten Wissens. In W. Schulz (Hrsg.), *Aspekte und Probleme der didaktischen Wissensstrukturierung* (S. 93-111). Frankfurt: Peter Lang.
- Eckert, A. (2000). Die Netzwerk-Elaborierungs-Technik (NET) – Ein computerunterstütztes Verfahren zur Diagnose komplexer Wissensstrukturen. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (S. 137-157). Göttingen: Hogrefe.
- Edelman, S. (1998). Representation is representation of similarities. *Behavioral and Brain Sciences*, 21, 449-498.
- Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie*. München: Reinhardt.
- Gower, J. C. & Dijksterhuis, G. B. (2004). *Procrustes problems*. Oxford: Oxford University Press.
- Inhelder, B., De Caprona, D. & Cornu-Wells, A. (Eds.). (1987). *Piaget Today*. Sussex: Erlbaum.
- Inhelder, B., Sinclair, H. & Bovet, M. (1974). *Apprentissage et structures de la connaissance*. Presses Universitaires de France.

- Janetzko, D. & Strube, G. (2000). Knowledge Tracking – Eine neue Methode zur Diagnose von Wissensstrukturen. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (S. 199-217). Göttingen: Hogrefe.
- Klauer, K. J. (1989). Allgemeine oder bereichsspezifische Transfereffekte eines Denktrainings. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 21, 185-200.
- Läge, D. (2001). *Ähnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Läge, D. (2004). *Konvergente und nicht-konvergente Veränderung der Ähnlichkeitsurteile beim Merkmalslernen*. AKZ-Forschungsbericht Nr. 2. Universität Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Daub, S., Bosia, L., Jäger, C. & Ryf, S. (2005). *Die Behandlung ausreißer-behafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag*. AKZ-Forschungsbericht Nr. 21. Zürich: Angewandte Kognitionspsychologie.
- Lee, K., Byatt, G. & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: testing the face-space framework. *Psychological Science*, 11, 379-385.
- Mandl, H. & Fischer, F. (Hrsg.). (2000). *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*. Göttingen: Hogrefe.
- Mandl, H. & Spada, H. (1988). *Wissenspsychologie*. München: Psychologie Verlags Union.
- Marx, W. & Hejj, A. (1989). *Subjektive Strukturen*. Göttingen: Hogrefe.
- Marx, W. & Läge, D. (1995). *Der ideologische Ring*. Göttingen: Hogrefe.
- Mathar, R. (1997). *Multidimensionale Skalierung*. Stuttgart: Teubner.
- Montada, L. (1995). Die geistige Entwicklung aus der Sicht Piagets. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (3. Aufl.) (S. 622-653). Weinheim: Psychologie Verlags Union.
- Piaget, J. & Inhelder, B. (1974). *Gedächtnis und Intelligenz*. Olten: Walter.
- Piaget, J. (1973). *Das Erwachen der Intelligenz beim Kinde*. Stuttgart: Klett.
- Piaget, J. (1976). *Die Äquilibration der kognitiven Strukturen*. Stuttgart: Klett.
- Piaget, J. (2001). *Studies in reflecting abstraction*. Sussex: Psychology Press.
- Plötzner, R., Beller, S. & Härder, J. (2000). Wissensvermittlung, tutoriell unterstützte Wissensanwendung und Wissensdiagnose mit Begriffsnetzen. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (S. 180-198). Göttingen: Hogrefe.
- Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 227-270). Cambridge, Mass.: MIT-Press.
- Scheele, B. & Groeben, N. (1984). *Die Heidelberger Struktur-Lege-Technik (SLT). Eine Dialog-Konsens-Methode zur Erhebung Subjektiver Theorien mittlerer Reichweite*. Weinheim: Beltz.
- Steiner, G. (1996). *Lernen – Zwanzig Szenarien aus dem Alltag* (2. vollst. überarb. Aufl.). Bern: Huber.
- Weber, S. & Schumann, M. (2000). Concept Mapping Software Tool (COMASOTO) zur Diagnose strukturellen Wissens. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (S. 158-179). Göttingen: Hogrefe.

3.

Der NMDS-Stresswert als Maß für Assimilation und Akkomodation beim Lernen von Wissensstrukturen

Zusammenfassung⁴

Assimilation und Akkomodation werden von Piaget als Teilprozesse bei der Bildung eines kognitiven Gleichgewichts im Wissenserwerb verstanden. Für den Bereich mathematisch-logischen und naturwissenschaftlich-funktionalen Wissens sowie für propositional-sprachliches Wissen ist eine solche Äquilibration hinlänglich belegt. Bislang fehlte jedoch ein geeignetes Verfahren, um den Prozess der Veränderung einer Wissensstruktur auch bei merkmalsorientiertem Faktenwissen messen zu können. Der Stresswert einer Nonmetrischen Multidimensionalen Skalierung ist – neben einer Reihe anderer Parameter – auch wesentlich durch die innere Konsistenz der zugrunde liegenden Datenmatrix beeinflusst. Aus diesem Grund kann er als Schätzwert für den Grad der Integriertheit des Wissens derjenigen Person dienen, von der diese Daten stammen. Zwei Lernexperimente, in denen die Versuchspersonen das hinzugewonnene Merkmalswissen jeweils bei Ähnlichkeitsurteilen zwischen den Objekten einzusetzen hatten, überprüfen anhand des NMDS-Stresswerts gezielt die Hypothesen, dass Assimilation mit einer Verringerung der Gesamtkonsistenz der Wissensstruktur einhergeht, während zeitlich darauf folgende Akkomodationsprozesse diese Konsistenz wiederherstellen. Die Resultate bestätigen für Assimilation und für Akkomodation eine Effektgröße, die den gegenläufigen Übungseffekt (gemessen bei nicht lernenden Kontrollgruppen) nicht nur ausgleicht, sondern signifikant übertrifft.

Schlüsselwörter: Assimilation, Akkomodation, Äquilibration, Wissenspsychologie, Sachwissen, Merkmalswissen, Wissenskarte, Kognitive Struktur, Kognitive Karte, Lernexperiment, Ähnlichkeitsurteil, Nonmetrische Multidimensionale Skalierung, Stresswert, Prokrustes-Transformation

3.1. Einleitung

Das Merkmalswissen einer Person über eine Anzahl von Objekten lässt sich mit so genannten „Kognitiven Karten“ darstellen (Läge, 2001, Marx & Hejj, 1989; Marx & Läge, 1995). Als dimensionale Gebilde zeigen sie die Relationen, die eine Person zwischen diesen Objekten sieht, in Form von kleineren oder größeren Distanzen auf und lassen als Gesamtbild damit Rückschlüsse auf die von der Person als wesentlich erachteten Merkmalsbereiche zu. Gleichzeitig erlauben Kognitive Karten, das Wissen über ein spezifisches Objekt durch deren Position in der Gesamtstruktur nachzuvollziehen.

Abbildung 18 gibt ein Beispiel für eine solche Wissenskarte. Sie wurde mittels Nonmetrischer Multidimensionaler Skalierung (NMDS) aus einer Matrix paarweiser Ähnlichkeitswerte generiert. In diesem Fall handelt es sich – um eine weitere Abbildung einzusparen – um die aus

⁴ Teile dieser Arbeit wurden am 43. Kongress der Deutschen Gesellschaft für Psychologie als Poster vorgestellt: Streule, R., Schlatter, K., & Läge, D. (2002). Assimilation und Akkomodation beim Wissenserwerb - gezeigt anhand Kognitiver Karten. In E. van der Meer, H. Hagendorf, R. Beyer, F. Krüger, A. Nuthmann, & S. Schulz (Hrsg.), *43. Kongress der Deutschen Gesellschaft für Psychologie 2002*. Lengerich: Pabst. S. 426.

den Merkmalsdaten von Experiment 2 errechnete Karte, aber man kann sich als Einführungsbeispiel auch vorstellen, dass diese NMDS-Karte auf der Basis eines Satzes individueller Ähnlichkeitsurteile entstanden ist. Durch Property Fitting (multiple lineare Regression; vgl. Läge, 2001) ließen sich im Nachhinein die drei Dimensionen „Bruttosozialprodukt“, „Landesfläche“ und „Einwohnerzahl“ mit genügend guten Korrelation in die Karte einpassen, so dass im Fall einer individuellen Wissenskarte interpretiert werden dürfte, dass die befragte Person bei ihren Urteilen vor allem auf diese Merkmale zurückgegriffen haben muss. Außerdem lässt sich bereits aus der Relation zu den anderen Objekten für ein beliebiges Land, z.B. für Frankreich, sagen, dass die Person hierunter ein relativ großflächiges, einwohnerreiches Land mit hohem Bruttosozialprodukt versteht.

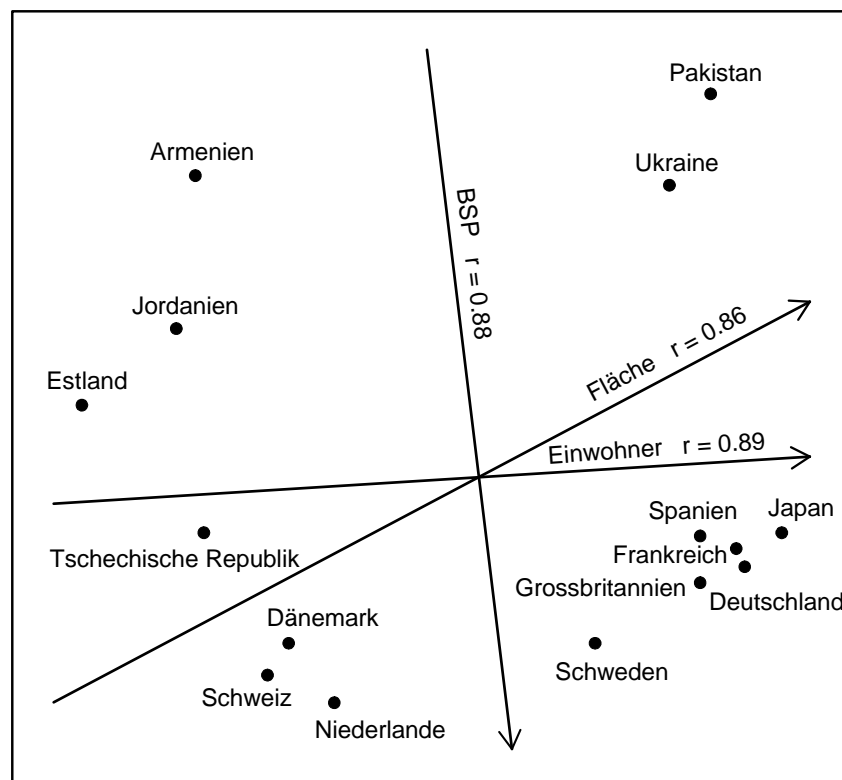


Abbildung 18. NMDS-Karte über 15 Länder und drei Merkmale, als Einführungsbeispiel und gleichzeitig als Zielstruktur des Lernens in Experiment 2.

Die Genese derartiger Wissenskarten im Laufe von Lernprozessen ist bislang überraschend schlecht untersucht. Lediglich Roth (2000) macht einen leider aber isoliert bleibenden Vorschlag, die Veränderung von Wissen im Sinne der Piaget'schen Theorie mit artifiziellen Neuronalen Netzwerken abzubilden. Neuerdings werden allerdings normative Stimmen laut, Wissensvermittlung stärker an die bestehenden Kognitiven Strukturen anzupassen (z.B. Novak, 2002, Bezug nehmend auf Ausubel, 2000). Dazu wäre aber eine systematische Modellierung dieser Strukturen auf individuellem Niveau notwendig, um über die rein deskriptive Ebene hinausgehen zu können, die beispielsweise Venville & Treagust (1998) in ihrer multimedial sehr sorgfältig aufgebauten Studie über konzeptuelle Veränderungen während eines Genetik-kurses anbieten. Anwendungen Nonmetrischer Multidimensionaler Skalierung (oder verwandter Methoden) in der Psychologie, Soziologie oder Pädagogik beschränken sich bis an-

hin jedoch fast ausschließlich auf die Untersuchung der Sichtweise von Populationen und argumentieren deswegen auf der Basis von Proximitätenmatrizen, die über eine Stichprobe gemittelt wurden. Gelegentliche Einzelfallstudien, wie z.B. eine über das Dinosaurierwissen eines Kindes (Chi & Koeske, 1983) oder über die geographische Vorstellung eines Herrn B. (Gigerenzer, 1981), bilden die Ausnahme. Aber selbst sie zeichnen nicht einen Lernprozess nach, sondern begnügen sich mit einer Momentaufnahme in Form einer Statusdiagnose.

Dabei sind derartige Wissenskarten aufgrund ihres relationalen Charakters eigentlich geeignet, die schrittweise Verbesserung von objektbezogenem Sachwissen auf Strukturebene nachzuzeichnen und dabei – etwa in der Tradition Piaget'schen Denkens – in der Terminologie von Assimilations- und Akkomodationsleistungen zu beschreiben, denn diese beiden (ursprünglich aus der Biologie entlehnten) Begriffe beschreiben ja kognitive Integrations- und Strukturierungsleistungen. Assimilation bedeutet Anpassung der Wirklichkeit oder der Umwelt an die eigene, aktuelle kognitive Organisation. Neue Information wird in bestehende Assimilationsschemata integriert. Akkomodation heißt Anpassung, Erweiterung oder Veränderung der kognitiven Schemata oder Strukturen in Richtung auf eine Angleichung an die Umweltanforderungen (für eine ausführliche Darstellung des Konzepts der kognitiven Adaptationen siehe Piaget, 1976).

Aus dieser Definition leitet Piaget seine zwei Grundpostulate ab, welche zentral für die Argumentationswege einer Gleichgewichtstheorie sind: „Erstes Postulat: Jedes Assimilationsschema hat die Tendenz zu wachsen, das heißt sich die Elemente einzuverleiben, die ihm äußerlich und mit seiner Natur verträglich sind. (...) Zweites Postulat: Jedes Assimilationsschema ist gezwungen, sich an die Elemente zu akkomodieren, die es assimiliert, das heißt sich entsprechend ihren Besonderheiten zu verändern, ohne deshalb seine Kontinuität (also seine Geschlossenheit als Zyklus interdependenter Prozesse) oder sein früheres Assimilationsvermögen zu verlieren...“ (Piaget, 1976, S. 14f.).

Das so entstehende Gleichgewicht (und auch die Probleme, die bei seiner Störung geschehen), sollte sich eigentlich in Wissenskarten gut detektieren lassen können, denn ihre anzunehmende Genese deckt sich praktisch eins zu eins mit den Definitionen und Postulaten Piagets: Eine Person wird neu gelernte Objekte so lange wie möglich in ihre bestehende Wissensstruktur integrieren. Erst wenn es durch die neuen Objekte (oder durch neues Merkmalswissen zu bereits bekannten Objekten) zu einem strukturellen Konflikt kommt, wird sie eine entsprechende Umstrukturierung ihrer kognitiven Wissensorganisation vornehmen. Dabei dürfte – ganz im Sinne des zweiten Postulates Piagets – die Kontinuität im Laufe des Prozesses erhalten bleiben.

Abbildung 19 zeigt ein Beispiel für einen Lerneffekt, der hauptsächlich auf einer Assimilationsleistung beruht: Fünf neue Länder sind hinzugekommen, welche in die bestehende Gesamtstruktur integriert werden konnten. Gleichzeitig ist diese Gesamtstruktur aber nicht ganz unverändert geblieben, sondern auch sie hat sich – um in Piagets Formulierung zu bleiben – „an die Elemente ... akkomodiert, die es assimiliert.“ Allerdings steht der Assimilationseffekt klar im Vordergrund, denn die Wissensstruktur hat ja keine nennenswerte qualitative Verän-

derung erfahren (der Abweichungswert aus der Prokrustes-Transformation ist gering), sondern wurde vielmehr erweitert.

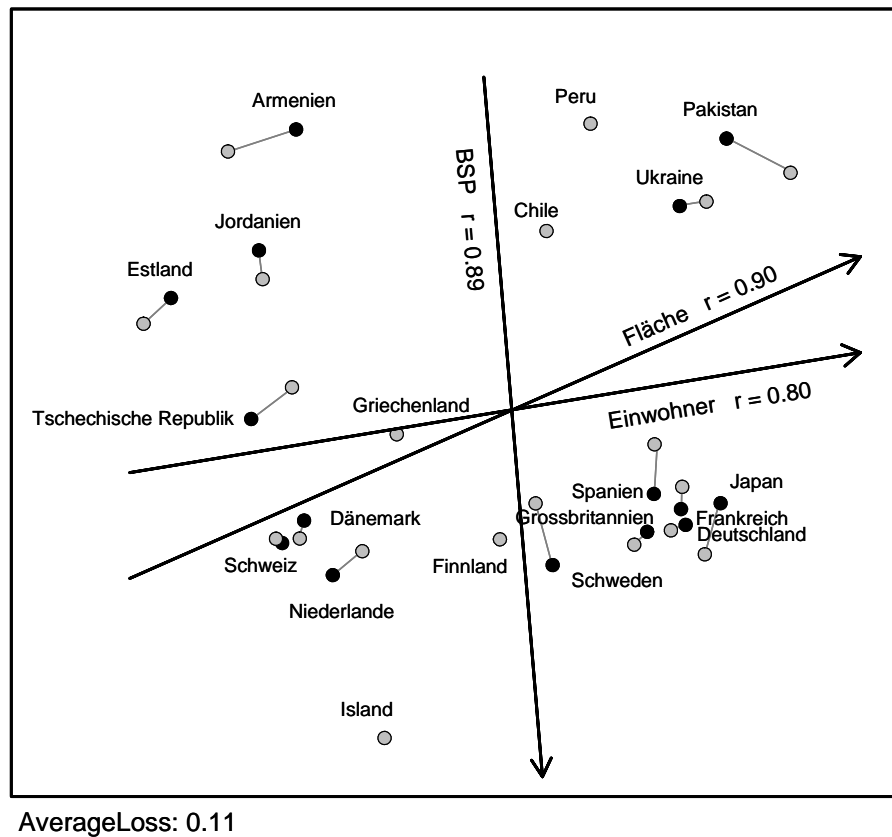


Abbildung 19. Erweiterung von Abbildung 1 um fünf weitere Länder (Chile, Finnland, Griechenland, Island, Peru). Die schwarzen Punkte stehen für die Normkarte aus Abbildung 1. Die grauen Punkte für die neu berechnete NMS-Lösung mit 20 Objekten. Die Abweichung der Positionen mit 15 Objekten verglichen mit den Positionen aus der Karte mit 20 Objekten beträgt AvgLoss = 0.11.

Anders sieht dies aus, wenn zu den bereits bekannten Objekten andere Merkmalsinformation berücksichtigt wird. Deren Integration kann die aus den Ähnlichkeitsurteilen resultierende Wissenskarte gehörig auf den Kopf stellen. Abbildung 20 zeigt ein solches Beispiel, in dem die Merkmale durchschnittliche Jahrestemperatur und durchschnittliche jährliche Niederschlagsmenge die Merkmale Fläche und Einwohnerzahl ersetzen. In den Worten von Piaget's Theorie hat hier ein Umbau der kognitiven Struktur im Sinne einer Akkomodation stattgefunden: Die relationalen Beziehungen haben sich neu organisiert, und in der Folge verändern die schwarz markierten Objekte ihre Position im Vergleich zur vorausgehenden Karte. Die Person hat gewissermaßen ihr „Weltbild“ hinsichtlich der Geographie geändert, nachdem sich die neu gelernten Merkmale nicht mehr mit dem alten Schema in Einklang bringen ließen. Dass nicht ein noch größerer Umbau entstanden ist, hängt vor allem an der konstanten Verwendung des Merkmals Bruttosozialprodukt in beiden Karten.

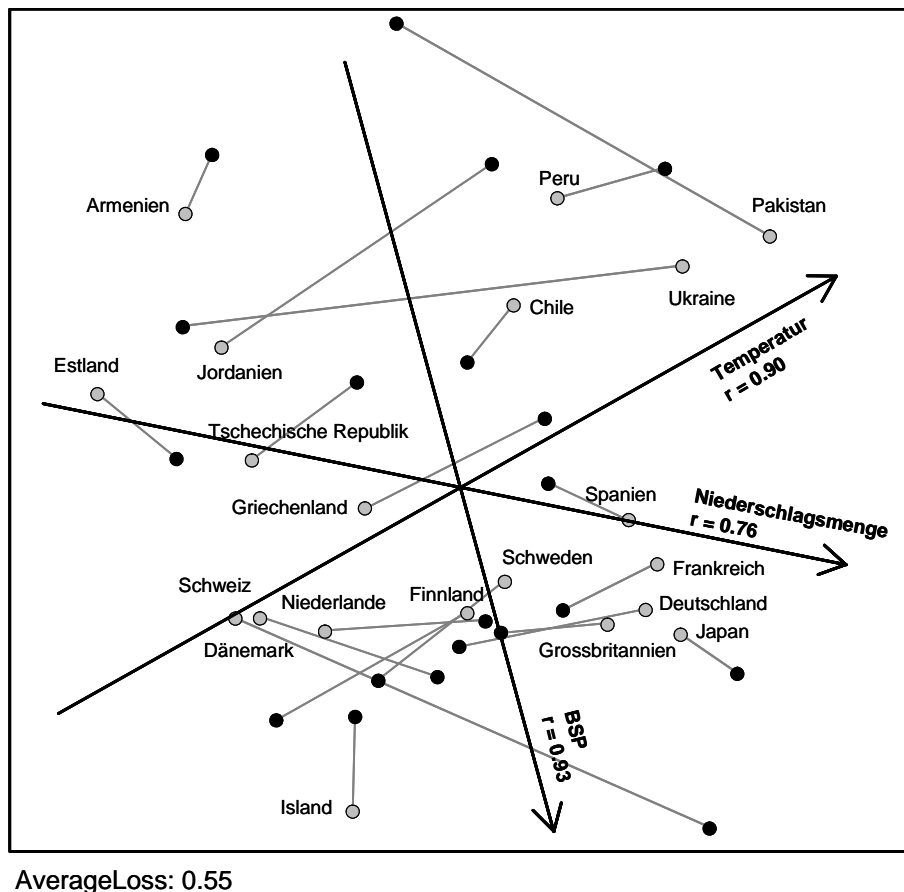


Abbildung 20. Veränderung der Karte im Sinne einer Akkomodation. Die grauen Punkte stehen für dieselben grauen Punkte in Abbildung 2. Die schwarze Punktekonfiguration basiert auf den Merkmalen BSP/Kopf, jährliche Durchschnittstemperatur und jährliche durchschnittliche Niederschlagsmenge. Ebenfalls sind die Regressionsgeraden für die schwarzen Punkte eingetragen. Deutlich zu erkennen ist der auf die unterschiedlichen Informationen rückführbare Umbau der Kartenstruktur mit einem AverageLoss von 0.55.

Die angeführten Beispiele sind idealtypisch und deswegen klar zu interpretieren. Sie zeigen anhand eines technischen Wertes (dem AverageLoss aus einer Prokrustes-Transformation) numerisch das Ausmaß der Veränderung der alten Struktur auf. Ist diese Veränderung für die ursprünglich bereits gekannten Objekte gering, so wurde das neue Wissen assimiliert, ist diese groß, so akkomodierte die Person. In der Realität gehen – wie Piaget betont – aber beide Prozesse Hand in Hand, und es dürfte schwierig sein, allein über einen (im Mittelbereich dann natürlich sehr unscharfen) Abweichungswert zwischen zwei Karten zwischen Akkomodation und Assimilation zu unterscheiden. Gleichwohl zeigen die Beispiele, dass es reizvoll wäre, für den Bereich solcher Wissenskarten ein geeignetes Messverfahren zu entwickeln. Gelingt das nämlich, dann könnte sich eine wesentliche Lücke schließen lassen, die bislang durch die starke Asymmetrie in der Erforschung regelbasierten Wissens auf der einen und merkmalsbasierten Sachwissens auf der anderen Seite besteht (Montada, 1995, liefert dazu einen hervorragenden Überblick). Diese Asymmetrie zugunsten des regelbasierten Wissens (für Felder wie Physik oder Mathematik liegt eine enorme Dichte aussagekräftiger Forschungsbefunde vor) dürfte zum einen dadurch geleitet sein, dass für die Messung physikalischer oder mathematischer Vorstellungen ein reichhaltiger Satz an Experimental- und Rechenaufgaben vorliegt, um das Wissen in jeder Altersgruppe zu untersuchen. Zum anderen mag aber auch eine

Rolle spielen, dass das Faktenwissen bislang in einer stark an der traditionellen, regelbasierten K.I.-Forschung orientierten Kognitionspsychologie nur als unverbundenes „lexikalisches Wissen“ angesehen wurde. Strube (1996) fasst in seinem Handbuchartikel zu *Wissen & Wissensrepräsentation* diese Sichtweise prägnant zusammen, wenn er folgerichtig von „dem primitivsten Format zur Wissensrepräsentation“ spricht (S. 806). Betrachtet man Merkmalswissen hingegen als in einem Wissensfeld integriert, dann lässt sich auch das Lernen von „trivialen“ Fakten in seiner Integration in das Gesamtwissen einer Person nachzeichnen.

Das ist aber leichter gefordert als getan. Das Betrachten einer einzelnen Punktkonfiguration allein führt nicht weiter. Der Vergleich zeitlich aufeinander folgender Wissenskarten (z.B. durch Prokrustes-Transformation) könnte ein Ansatz sein. Doch dieser Vergleich einer Konfiguration mit dem vorherigen Wissensstand lässt assimilative oder akkomodative Aspekte wohl nur im Extremfall (wie in den beiden gewählten Beispielen) klar zu Tage treten. Eine Kognitive Karte besteht ja nur aus Relationen der Objekte zueinander, und so fehlt der archimedische Punkt, an dem sich der Unterschied zwischen assimilativer Einpassung von neuem relationalen Wissen in die alte Struktur und globaler Veränderung der alten Struktur durch Akkomodationsleistungen festmachen ließe.

Wir schlagen deswegen ein Maß für die innere Konsistenz abgegebener Ähnlichkeitsurteile vor. Ganz im Sinne Piagets gehen wir nämlich davon aus, dass neues Wissen solange assimiliert wird, solange dadurch das bestehende Weltbild als ganzes in sich stimmig bleibt. In seinem zweiten Postulat bezeichnet Piaget dieses jedoch als einen Grenzfall, wenn er davon spricht, dass ein Assimilationsschema immer auch zu akkomodieren habe. Assimilation als Prozess erzeugt damit also so etwas wie „Reibung“ beim schon Bestehenden: kleine Inkonsistenzen treten zu Tage, werden aber solange es geht toleriert. Akkomodation hingegen bewirkt das Auflösen dieser Inkonsistenzen durch Restrukturierung des Gesamtfeldes.

Beziehen wir diesen Gedanken auf Ähnlichkeitsurteile, so ergibt sich folgendes: Befragen wir eine Person unmittelbar nach einem Lerndurchgang, so sollte sie versuchen, die neu gelernte Merkmalsinformation zusätzlich zum bisherigen Wissen in jedes dieser Urteile einfließen zu lassen. In der Summe sollte das dann zu einer qualitativ verbesserten Wissenskarte führen (gegeben einmal, die gelernte Merkmalsinformation ist funktional und sachrichtig). Gleichzeitig sollte aber eine gewisse Beharrungstendenz hinsichtlich des Alten bestehen bleiben. Die einzelnen Ähnlichkeitsurteile werden dadurch zwar in der Summe qualitativ besser, verlieren jedoch an innerer Konsistenz. Assimilation als Integration des neuen Wissens in die alte Struktur erzeugt damit die oben beschriebene „Reibung“. Erst ein gegenläufiger Prozess, nämlich die Integration des alten in die strukturbildenden Fakten des neuen Wissens (auch so ließe sich „Akkomodation“ beschreiben) löst diese „Reibung“ auf. Ähnlichkeitsurteile sollten nach einem Akkomodationsprozess in sich stimmiger sein als vorher.

Wann findet Assimilation statt, wann Akkomodation? Laut Piaget wird ein Gleichgewicht hergestellt durch die Elaboration bestehender Strukturen und deren Vergleich mit den neu hinzugekommenen Informationen (Piaget, 1976, S. 167). Daraus folgt, dass diese regulierenden Kompensationen kognitive Ressourcen beanspruchen, welche abhängig von den entstehenden Widerständen/Störungen zwischen den zuvor neu erworbenen Informationen und den

alten Strukturen unterschiedliche zeitliche Beanspruchungen erfordern. Zeitlich gesehen geht Assimilation der Akkomodation damit stets voraus: Zuerst wird eine neue Information an bestehende Strukturen angehängt; im Fall des Merkmalslernens werden die dem Wissensraum zugrunde liegenden Merkmalsdimensionen mit Werten belegt. Das Faktenwissen gewinnt somit an Qualität. Eine kognitive Struktur oder ein Schema in einem Gleichgewichtszustand zeichnet sich aber darüber hinaus durch eine in sich stabile Relationalität zwischen den Merkmalsausprägungen der Objekte aus. Eine Person kann zwar über objektiv richtiges Faktenwissen auf der Ebene von Einzelinformationen verfügen, es gelingt ihr aber nicht, die Beziehungen dieser Fakten zu anderen Inhalten konsistent zu strukturieren und in kongruente Beziehungen zu setzen. Dieser Zustand ist derjenige kurz nach dem Lernen von neuem Wissen. Die Neuaufnahme von Wissensinhalten erfordert immer eine Restrukturierung, eine Regulierung der Beziehungen zwischen den Objekten, da eben eine Erweiterung einer kognitiven Struktur nicht nur die alleinige Anbindung beinhaltet, sondern zudem ein Gesamtbild konstruiert werden muss. Dieser akkomodative Aspekt der Verarbeitung setzt nach diesen Überlegungen später ein und erfordert vermutlich auch mehr Zeit.

Assimilation von Wissen bedeutet also Verringerung der Konsistenz bei gleichzeitiger qualitativer Verbesserung der Gesamtstruktur. Akkomodation ist der nachfolgende Prozess der Verringerung der Inkonsistenzen.

Ein Maß für die innere Konsistenz von Ähnlichkeitsurteilen ist deren geometrische Umsetzbarkeit in eine Karte gegebener Dimensionalität. Die Nonmetrische Multidimensionale Skalierung gibt für jede Lösung den Stresswert an, also denjenigen Anteil bezogen auf die Gesamtheit aller Distanzen, um den einzelne Distanzen in der Konfiguration hinsichtlich der gegebenen Ordnung in den Daten zu groß oder zu klein ausfallen. Abbildung 21 gibt ein Beispiel dafür, wie sich der Stresswert aus einzelnen Einpassungsfehlern zusammensetzt. Die grundsätzliche Idee, den Stresswert für jeden Punkt der Konfiguration zu betrachten, hat bereits Borg, 1978 formuliert). Dieser Stresswert ist ein wichtiges technisches Maß der NMDS, denn die iterativen Approximationsalgorithmen sind so gebaut, dass sie beim Finden der endgültigen Lösung diesen Abweichungswert zwischen Daten und Resultat minimieren (vgl. Borg & Groenen, 1997, für die Logik des Verfahrens und zur mathematischen Herleitung Mathar, 1997). Außerdem gibt der Stresswert auch einen Hinweis darauf, ob die skalierten Daten überhaupt Struktur enthalten (für eine Monte-Carlo-Studie zur zufallskritischen Absicherung des Stresswertes siehe Gigerenzer, 1981).

In diesen Stresswert fließen eine Reihe von Einflussgrößen ein: Neben der Dimensionalität der Lösung und der Anzahl der zu skalierenden Objekte auch die Wahl der Ausgangskonfiguration, die Art der Behandlung von Rangbindungen in den Daten, das gewählte Verfahren zur Berechnung von Zieldistanzen, der gewählte Verschiebealgorithmus, die Stärke der resultierenden Clusterung der Lösung, die Konsistenz der Proximitäten für die „Eckpunkte“ der Lösung und vieles mehr. Die innere Konsistenz der zu skalierenden Daten ist dabei zwar ein wichtiger, aber eben nur ein Einflussfaktor unter vielen.

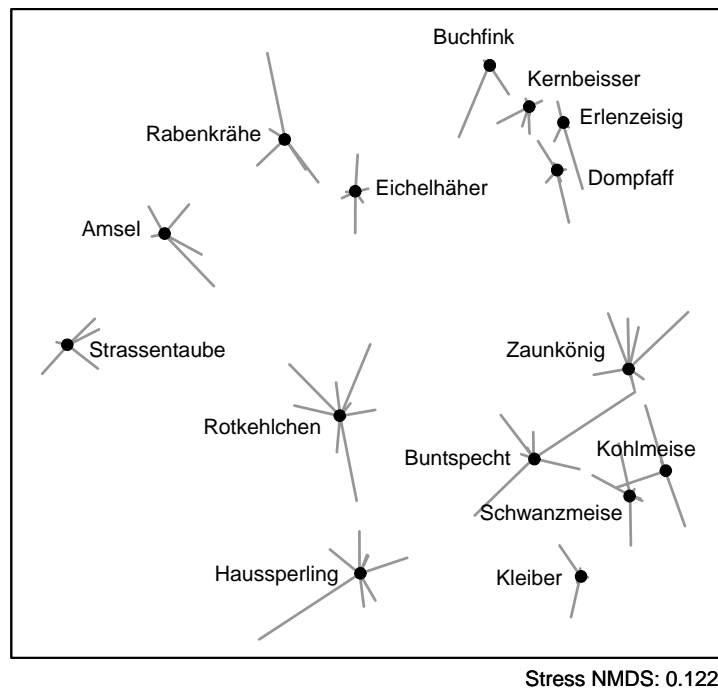


Abbildung 21. Eine individuelle Karte aus Experiment 1 mit eingezeichnetem Distanzstress (symmetrischer Stress pro Objekt = graue Striche) und Gesamtstress = 0.12.

NMDS-Algorithmen sind normalerweise so gebaut, dass sie als einziges Ziel den Gesamtstress minimieren. Die Frage der inneren Konsistenz der Daten spielt dabei nur eine untergeordnete Rolle. Lediglich ein robuster Algorithmus würde versuchen, in den Daten das „Signal“ (das heißt den zueinander passenden Teil der Proximitäten) zu entdecken und möglichst fehlerfrei darzustellen. Was an Inkonsistenzen übrig bleibt, geht dann als „Rauschen“ (und bei groben Inkonsistenzen gegebenenfalls als Ausreißer) in den Stresswert ein. Mit RobuScal (Läge, 2001) liegt ein solch robuster Algorithmus vor. Er gibt Anlass zu der Hoffnung, dass über Stresswertunterschiede von NMDS-Lösungen auch eine verlässliche Aussage über die innere Konsistenz einer Proximitätenmatrix möglich ist (Läge, Daub, Bosia, Jäger & Ryf, 2005). Angenommen, es gäbe einen perfekt robusten NMDS-Algorithmus, der es erlauben würde, Strukturanteil und Rauschen in den Ähnlichkeitsurteilen exakt auseinanderzudividieren (ein Wunschtraum!), dann würde ein assimilativer Prozess ein erhöhtes Rauschen und ein akkomodativer Prozess reduziertes Rauschen bedeuten. Leider kann man sich dieser Wunschvorstellung nach perfekter Zerlegung einer Ähnlichkeitsmatrix in Struktur und Rauschen mathematisch nur unvollkommen annähern, auch mit einem (hinsichtlich der anderen existierenden Algorithmen) so robusten Verfahren wie RobuScal. Trotzdem soll im Folgenden versucht werden, über den Stresswert der NMDS-Lösungen eine Schätzung für Akkomodation und Assimilation vorzunehmen.

Aus lerntheoretischer Sicht wäre die Verbindung der Überlegungen in der Denktradition Piagets mit der technischen Aspekten der Skalierung folgende: Unmittelbar nach einer Lerneinheit wird eine Person versuchen, das gerade erlernte Merkmalswissen in Verbindung zum bestehenden Wissen zu bringen, ohne das Bestehende stärker anzutasten als nötig. Sie hat zum Zeitpunkt der auf das Lernen folgenden Ähnlichkeitsurteile noch keine Gelegenheit gehabt, sich grundsätzlichere Gedanken über einen notwendigen Umbau der Gesamtstruktur zu

machen. Deswegen sind unmittelbar nach dem Dazulernen von neuer Information zunächst einmal assimilative Prozesse zu erwarten. Diese erhöhen das Rauschen in der Ähnlichkeitsmatrix, was sich in einer robusten NMDS als erhöhter Stresswert auswirkt – selbst wenn sich inhaltlich die Lösung verbessert.

Nach einer gewissen Zeit (in den präsentierten Experimenten wird das jeweils eine Woche sein) hat sich das neue Wissen „gesetzt“ (und ein Teil davon wurde vielleicht auch vergessen). Wenn die Person jetzt Ähnlichkeitsurteile abgeben soll, dann hat sie weniger Mühe, altes und neues Wissen zusammenzubringen. Das integrierte neue Wissen wird sich konsistenter auf alle Ähnlichkeitsurteile auswirken, und das wird das Rauschen in der Urteilmatrix senken. Eine robuste NMDS produziert daraus eine Karte mit geringerem Stresswert, selbst wenn diese nicht unbedingt inhaltlich besser (d.h. näher an einem Normmodell gelegen) ist als die Karte vor einer Woche (ein Teil der Fakten wird ja vermutlich inzwischen vergessen sein). Das nach einer Woche resultierende geringere Rauschen würde für Umbau der Struktur in Richtung auf ein konsistentes Gesamtbild, also für Akkomodation sprechen.

Vorhergesagt wird also eine Erhöhung des Stresswertes unmittelbar nach dem Lernen (bei gleichzeitiger qualitativer Verbesserung der Wissenskarte) und eine Verringerung nach einer Woche Latenz. Wegen der vielen Faktoren, die auf den Stresswert einer NMDS-Lösung einwirken, ist hier die Wahrscheinlichkeit besonders groß, einen Beta-Fehler zu begehen, d.h. aufgrund der vielen intermittierenden Einflüsse auf den Stresswert einen Effekt von Assimilation und Akkomodation nicht zu finden, obwohl er eigentlich da wäre. Anders gewendet: Sollte sich der vorhergesagte Effekt finden lassen, dann ist er eher verborgen im Gestrüpp zu erwarten als auf dem Präsentierteller.

Gleichzeitig sollte man bei Vorhersagen hinsichtlich der Stresswerte auch berücksichtigen, was üblicherweise bei mehrfach hintereinander abgegebenen Ähnlichkeitsurteilen geschehen sollte (wenn nicht gelernt wird): Liegt wenig Zeit zwischen den Urteilen, dann sollten sich die herangezogenen Kriterien verfestigen. Zu erwarten ist also eine Art Übungseffekt, der mit einer Erhöhung der Urteilskonsistenz und somit mit einer Verringerung des Stresswertes einhergeht. (Qualitativ sollte sich eine Karte dadurch natürlich nicht verbessern, sie sollte nur konsistenter werden.) Liegt hingegen viel Zeit zwischen der Abgabe von Ähnlichkeitsurteilen, dann sollte sich dieser Übungseffekt nicht einstellen. Hat es ihn zuvor gegeben, dann lässt sich sogar erwarten, dass er nach einiger Zeit wieder ganz oder teilweise verschwindet.

Diese Annahmen hinsichtlich des Abgebens von Ähnlichkeitsurteilen bei gleich bleibendem Wissen kontrastieren scharf zu denen bei Assimilation und Akkomodation. Da sie genau gegenläufig sind, und da ein Übungseffekt im Abgeben von Ähnlichkeitsurteilen natürlich auch dann eintritt, wenn dazugelernt wird, könnte dieser Übungseffekt den erwarteten Effekt durch Assimilation und Akkomodation auch auslöschen. In jedem Fall muss letzterer Effekt größer sein als der Übungseffekt, um in den Resultaten sichtbar werden zu können.

Im Folgenden werden dazu zwei Lernexperimente präsentiert. Das erste basiert auf recht anspruchsvollem, von den Merkmalsinformationen her nicht so leicht zu integrierendem Wissen (Ornithologie). Im zweiten Experiment ist das Lernmaterial wesentlich einfacher gestrickt (drei intervallskalierte Merkmale zu 15 Ländern). In beiden Experimenten sollten sich für

Assimilations- und Akkomodationsprozesse aber vergleichbare Befunde ergeben. Das Lernmaterial ist jeweils so konstruiert, dass auch eine Personengruppe ohne spezielles Lernprogramm mit ganz normalem Vorwissen bereits halbwegs vernünftig strukturierte Ähnlichkeitsurteile abzugeben in der Lage sein sollte (damit der angesprochene Übungseffekt gemessen werden kann).

3.2. Methode

3.2.1. Experiment 1

Aufbau des Experiments. In einer Längsschnittuntersuchung hatten die Versuchspersonen die Aufgabe, 15 Vogelarten anhand vier charakteristischer Merkmalsbereiche zu lernen. In drei Sitzungen, die jeweils im Wochentakt stattfanden, erhielten sie dazu Informationen zu diesen 15 Objekten. Mittels paarweiser Ähnlichkeitsurteile wurde dann die Auswirkung der Lerneinheiten auf die individuellen Wissenskarten der Versuchspersonen überprüft.

Die Versuchspersonen wurden vor dem Experiment zufällig in drei Gruppen aufgeteilt: zwei Experimentalgruppen mit unterschiedlichem Treatment (Exkursionslerner und Kärtchenlerner) und eine Kontrollgruppe. Die Kärtchenlerner erhielten in allen drei Lernphasen Kärtchen der 15 Vogelarten mit merkmalsstrukturierten Informationen. Die Exkursionslerner nahmen im ersten Lerndurchgang mit einem Ornithologieexperten an einer Exkursion teil, an der sie alle 15 Arten beobachten konnten. Die beiden folgenden Lernphasen unterschieden sich nicht von denen der Kärtchenlerner. Die Kontrollgruppe durchlief nur die Wissensdiagnose, nahm jedoch nicht an den Lerneinheiten teil. Der Aufbau des Experiments ist in Abbildung 22 skizziert.

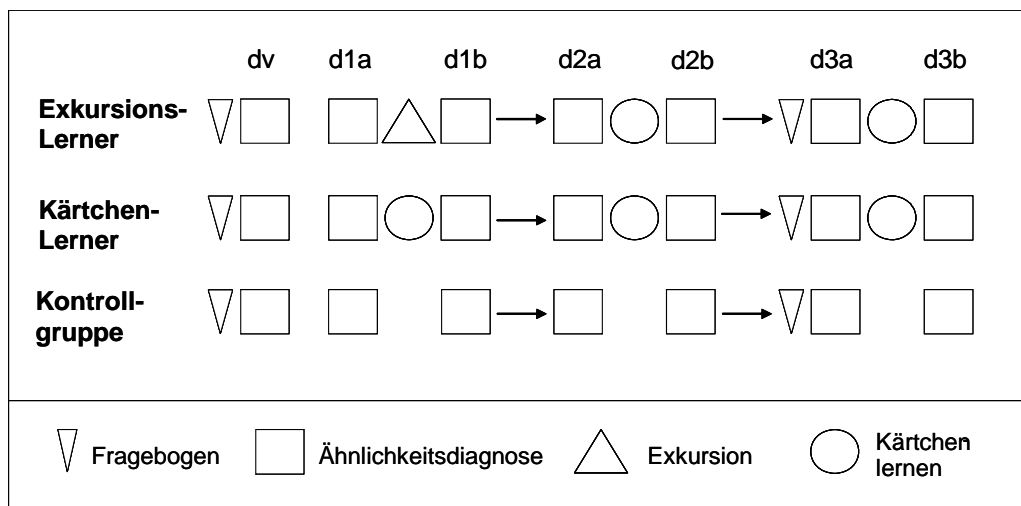


Abbildung 22. Schematische Darstellung des Ablaufs von Experiment 1. Die erste Ähnlichkeitsdiagnose in Experiment 1 wird im Resultateteil nicht berichtet, die zweite Diagnose mit d1a (Diagnose der ersten Sitzung vor dem Lernen, die dritte Diagnose mit d1b (= Diagnose der ersten Sitzung nach Lernen), die vierte Diagnose mit d2a (= Diagnose der zweiten Sitzung vor Lernen) usw. markiert sein.

Vor der ersten und vor der dritten Lernphase durchliefen alle Versuchspersonen einen merkmalsorientierten Sachwissenstest, indem ein Multiple-Choice-Fragebogen zu allen 15 Vogelarten die vier Merkmalsbereiche aus den Lerneinheiten testete. (Der zweite Test fand mit Absicht nicht unmittelbar nach einer Lerneinheit statt, um Kurzzeiteffekte zu vermeiden. Vor der dritten Lerneinheit lag eine Woche lernfreie Zeit, so dass nur der langfristige Erfolg aus den beiden ersten Durchgängen gemessen wurde.) Vor und nach jeder Lerneinheit wurde der aktuelle Wissensstand aller Versuchspersonen mittels paarweiser Ähnlichkeitsurteile auf der Strukturebene erhoben. Vor der ersten Lerneinheit gaben die Versuchspersonen den Satz von 105 Ähnlichkeitsurteilen zweimal ab. Damit lassen sich allfällige Übungseffekte beim Umgang mit diesem Verfahren zeigen (Der zweite der beiden Datensätze (d1a) wird dann bei der Auswertung als Messung des Vorwissens angesehen. Der erste Datensatz (dv) wird nicht berichtet, da keine Effekte aufgetreten sind). Die Kontrollgruppe durchlief dieselbe Wissensdiagnostik.

Auswahl der Vogelarten und Bestimmung der Zielstruktur. Die Auswahl der 15 Vogelarten, so wie sie sich im Expertenmodell (Abbildung 6) präsentiert, hatte einigen Kriterien zu folgen. (1) Für die Exkursion mussten alle Arten im Raum Zürich im selben Habitat leben und dort leicht zu finden sein, und zwar auch im Winter, als dieses Experiment durchgeführt wurde. (2) Einzelne Arten sollten nicht dermaßen von allen anderen Vögeln verschieden sein, dass auswertungstechnisch betrachtet die Gefahr degenerierter NMDS-Lösungen bestanden hätte (was zu unrealistisch niedrigen Stresswerten führen würde). Auch sollten die Arten nicht nur wenigen, scharf getrennten Verwandtschaftsgruppen angehören, um übermäßig kategoriale Lösungen zu vermeiden. Das war technisch wichtig, weil sich sonst ein möglicher Lernfortschritt nicht numerisch fair durch Prokrustes-Transformation von NMDS-Karten messen lassen würde. (3) Ca. sechs Arten sollten Laien (zumindest oberflächlich) bekannt sein, damit eine gewisse Grundstruktur vorhanden war. Mindestens sechs Arten sollten aber den Versuchspersonen völlig unbekannt sein, um ausreichenden Spielraum für das Lernen zu besitzen.

Die ersten beiden Restriktionen führten zu einer eingegrenzten Liste möglicher Arten. Diese wurden in einem Vorversuch 30 Studierenden der Universität Zürich in einem Fragebogen vorgelegt, um Bekanntheitsgrad und Vorwissen einschätzen zu können. Daraus erfolgte dann die Auswahl der 15 am geeignetsten erscheinenden Arten.

Auf Strukturebene benötigen wir eine Wissensnorm, mit der jede Lernerkarte zu jedem Zeitpunkt evaluiert werden kann. Da diese Karte für die numerische Bestimmung des Lernerfolgs sehr wichtig ist, wurde große Mühe darauf verwendet, sie so akkurat wie möglich zu erheben. Deshalb wurden die 15 ausgewählten Arten acht erfahrenen Ornithologie-Experten vorgelegt, die jeweils die paarweisen Ähnlichkeiten einzuschätzen hatten. Bei vier von diesen Experten konvergierten die resultierenden NMDS-Karten in sehr hohem Masse, so dass ihre gemittelten Urteile mit einem Gewicht von 50% zur Grundlage für die Normkarte herangezogen wurden. Um sicherzustellen, dass das Normmodell auch wirklich die zu lernenden Merkmalsbereiche abdeckte, wurden die individuellen Ähnlichkeitsurteile nach dem letzten Lernen für jede der beiden Treatment-Gruppen gemittelt. Jede dieser beiden Ähnlichkeitsmatrizen floss mit einem Gewicht von 25% in die durch NMDS zu bildende Normkarte ein (Abbildung 23).

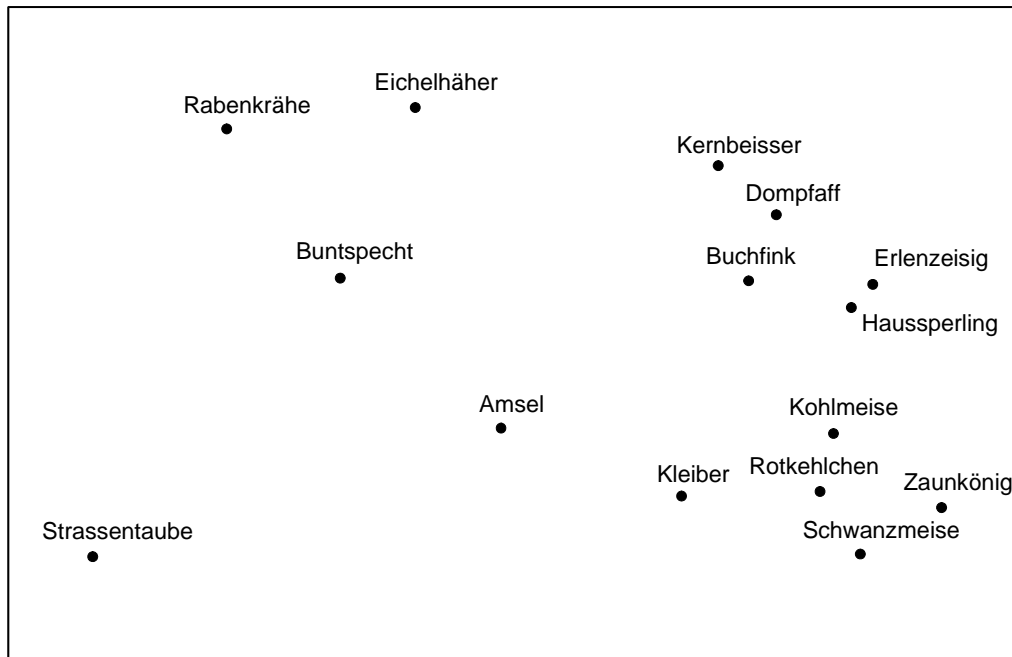


Abbildung 23. NMDS-Karte, die in Experiment 1 die Zielstruktur des Lernens ergibt.

Lernmaterialien. Die Exkursion fand in zwei Sechsergruppen statt und dauerte ca. eineinhalb Stunden. Die Lerner konnten alle 15 Arten unter Anleitung und begleitenden Kommentaren eines erfahrenen Exkursionsleiters per Fernglas beobachten. Zusätzlich zu seinen Auskünften hatten die Exkursionsteilnehmer die Lernkärtchen in der Hand und konnten sich bei Bedarf daran orientieren.

Diese Lernkarten setzten sich je Vogelart aus drei Teilen zusammen: aus a) einer farbigen Abbildung der Vogelart, aus b) einem kurzen, beschreibenden Text, der neben Besonderheiten und leicht eingänglichen Kuriositäten über die Art eine Merkmalsbeschreibung zu Lebensraum, Nahrung und Nistgewohnheiten enthielt (so wie man sie in Standard-Vogelbüchern findet), sowie aus c) einer tabellarischen Merkmalsliste, die alle wichtigen Merkmale aus dem beschreibenden Text tabellarisch aufführte.

In den individuell durchgeführten Lerndurchgängen erhielten die Versuchspersonen jeweils dieselben Lernmaterialien wieder vorgelegt. Sie waren in der Organisation ihres Lernens weitgehend frei, hatten sich aber in den beiden ersten Sitzungen jeweils 30 Minuten und in der dritten Sitzung 20 Minuten damit zu beschäftigen. Als Kriterien wurde ihnen an die Hand gegeben, dass sie versuchen sollten, aufgrund der Abbildungen die Arten zu benennen und dass sie in der verbleibenden Zeit alle Kärtchen mindestens einmal durchgelesen haben sollten. Sie wussten, dass sie Ähnlichkeitsurteile über diese Arten abzugeben hätten und konnten bereits beim Lernen die Kärtchen strukturiert auf einem Tisch ausbreiten.

Versuchspersonen. An der Hauptuntersuchung nahmen 31 Studierende der Universität Zürich im Alter von 22–27 Jahren teil. Sie besaßen keine spezifischen ornithologischen Vorkenntnisse. In den beiden Experimentalgruppen befanden sich jeweils 4 Frauen und 8 Männer, die Kontrollgruppe bestand aus 5 Frauen und 4 Männern. Die Gruppenzuteilung erfolgte vor dem Experiment und zufällig.

3.2.2. Experiment 2

Aufbau des Experiments. In einer Längsschnittuntersuchung hatten die Versuchspersonen die Aufgabe, zu 15 Ländern die drei Merkmale Landesfläche, Einwohnerzahl und Bruttosozialprodukt so zu lernen, dass sie daraus eine Wissenskarte aufbauen konnten. In vier Sitzungen, die jeweils im Wochentakt stattfanden, erhielten sie dazu die relevanten Informationen zu diesen 15 Objekten. Mittels paarweiser Ähnlichkeitsurteile unmittelbar vor und unmittelbar nach der Lerneinheit wurde dann die Auswirkung der Lerneinheiten auf die individuellen Wissenskarten der Versuchspersonen überprüft. Zu Beginn jeder Sitzung erhob ein computerisierter Multiple-Choice-Fragebogen den aktuellen Wissensstand über die drei Merkmale zu jedem Land. In diesem Fragebogen hatten die Versuchspersonen die Länder einer Liste aller einzelnen Merkmalsausprägungen (plus Distraktoren) zuzuordnen.

Die Versuchspersonen wurden vor dem Experiment zufällig in eine Experimentalgruppe und eine Kontrollgruppe aufgeteilt. Die Versuchspersonen der Kontrollgruppe lernten nicht, sondern führten lediglich das komplette Programm der Wissensdiagnostik durch. Der Aufbau des Experiments ist in Abbildung 24 skizziert.

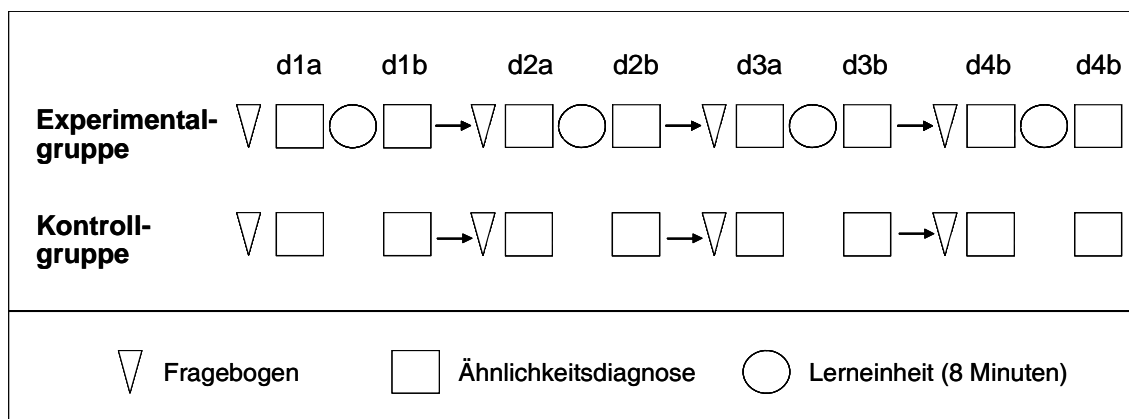


Abbildung 24. Schematische Darstellung des Ablaufs von Experiment 2.

Auswahl der Länder und Bestimmung der Zielstruktur. In einem Vorversuch mit insgesamt 46 Versuchspersonen hatte sich aus einer Vorauswahl von 26 Ländern der Pool von 15 Ländern als gut geeignet herausgestellt, der zur Grundlage der Zielstruktur gemacht wurde. Außerdem hatten sich die drei intervallskalierten Merkmale Landesfläche, Einwohnerzahl und Bruttosozialprodukt pro Kopf als günstig für Lernen, Wissensdiagnose und Zielstruktur erwiesen. Die Zielstruktur errechnete sich über die Unähnlichkeiten der einzelnen Länder auf diesen drei Merkmalen. Dazu wurde folgende Formel verwendet:

$$U\ddot{A}_{(x,y)} = \frac{|(x_i - y_i) / (x_i + y_i)| + |(x_j - y_j) / (x_j + y_j)| + |(x_k - y_k) / (x_k + y_k)|}{3}$$

Wobei: $U\ddot{A}_{(x,y)}$: Unähnlichkeitsquotient eines Länderpaares
 x, y : Länderindex
 i : Einwohnerzahl
 j : Fläche in qkm
 k : BSP pro Kopf in US\$

Dieser Quotient $U\ddot{A}_{(x,y)}$ beschreibt den Betrag des prozentualen Unterschiedes je zweier Länder aufgrund ihrer Merkmalsausprägungen. Die resultierende Dreiecksmatrix wurde mittels NMDS (RobuScal-Algorithmus; Läge et al., 2005) in eine zweidimensionale Karte transformiert, die für die Qualitätsabschätzung der individuellen Wissenskarten jeweils als Norm gilt.

Lernmaterialien. Das Lernmaterial bestand aus 15 Kärtchen, auf denen zu je einem Land die Angaben über Einwohnerzahl, Fläche und Bruttosozialprodukt standen. Die Versuchspersonen waren frei in ihrer Lernstrategie; sie konnten die Lernkarten ordnen und Gruppen bilden, ausbreiten usw. Eine Lerneinheit dauerte jeweils acht Minuten.

Versuchspersonen. 37 Personen (17 weiblich; 20 männlich) im Alter zwischen 19 und 33 Jahren nahmen an dem Experiment teil. Es handelte sich dabei hauptsächlich um Studentinnen und Studenten der Psychologie, Soziologie, Tiermedizin und Pharmazie der Universität Zürich. Hinzu kamen einige Versuchspersonen ohne Hochschulausbildung. 25 Personen wurden der Experimentalgruppe zugeordnet, 12 der Kontrollgruppe. Die Verteilung auf die beiden Stichproben erfolgte im Einzelfall zufällig, insgesamt gesehen aber so, dass die Variablen Alter, Geschlecht und Schulbildung in etwa gleichverteilt waren.

3.2.3. Auswertung der Daten und erwartete Befunde

Auswertung der Fragebögen: erwartete Verbesserung des Merkmalswissens. Voraussetzung für Assimilations- und Akkomodationsleistungen auf der Ebene der Wissenskarte ist die Verbesserung des Merkmalswissens. Deswegen werden zunächst die Fragebögen ausgewertet. Die Versuchspersonen der Lerngruppen sollten sich signifikant verbessert haben, die Versuchspersonen der Kontrollgruppen nicht.

Auswertung jeder einzelnen Ähnlichkeitsmatrix. Jede Ähnlichkeitsmatrix jeder Versuchsperson zu jedem Diagnosezeitpunkt wurde mit dem RobuScal-Algorithmus in eine zweidimensionale NMDS-Karte überführt. Diese Karte wurde dann mittels Prokrustes-Transformation auf das Normmodell für das jeweilige Wissensgebiet gelegt. Als numerisches Resultat für die Wissensgüte diente der AverageLoss, als Maß für die Konsistenz der Karte der Stresswert aus der NMDS.

Verlauf der Qualität der Wissenskarten. Für die Wissenskarten wird erwartet, dass die AverageLosses in der Prokrustes-Transformation jeweils nach einer Lerneinheit kleiner sind als in der Karte vor dieser Lerneinheit. Außerdem wird angenommen, dass in der Woche nach dem Lernen ein Teil des Zugewinns durch Vergessen wieder verloren geht. Die Wissenskarten der Personen aus den Kontrollgruppen sollten sich nicht systematisch verbessern. Am Ende des

Experiments sollten die Versuchspersonen der Experimentalgruppen die klar besseren Karten aufweisen (Diese Hypothesen sind die minimal aufzustellenden, um von „Lernen“ auf Strukturebene sprechen zu können – der notwendigen Voraussetzung, dass es sich überhaupt lohnt, nach Assimilations- und Akkomodationsprozessen bei der Veränderung der Strukturen zu suchen. Eine wesentlich detailliertere Analyse der Verbesserung der individuellen Wissenskarten beim Merkmalslernen folgt einer völlig anderen Fragestellung als der hier präsentierten und wird bei Kapitel 2 in dieser Arbeit behandelt.).

Übungseffekt bei wiederholter Abgabe von Ähnlichkeitsurteilen. Für die Personen der Kontrollgruppen wird erwartet, dass sie bei wiederholter Abgabe von Ähnlichkeitsurteilen innerhalb einer Sitzung konsistenter werden, was sich in einem systematisch niedrigeren Stresswert der individuellen Wissenskarten äußert. Dieser Übungseffekt sollte in der einwöchigen Pause zwischen den Sitzungen ganz oder teilweise wieder verschwinden.

Erhöhung der Inkonsistenzen durch Assimilationsprozesse. Für die Versuchspersonen der Experimentalgruppen erwarten wir, dass die unmittelbar nach dem Lernen erhobenen Karten zwar inhaltlich besser, dafür strukturell aber inkonsistenter sind als vor dem Lernen. Sie sollten damit einen erhöhten Stresswert aufweisen.

Verringerung der Inkonsistenzen in den Lernpausen. In der Zeit zwischen zwei Sitzungen sollte sich das dazu gewonnene Wissen strukturell „setzen“. Wir erwarten, dass die Wissenskarten der Versuchspersonen in der Ähnlichkeitsdiagnose nach einer Woche Pause durch diese Akkomodationsprozesse niedrigere Stresswerte aufweisen als unmittelbar nach dem vorausgegangenen Lernen.

3.3. Resultate

Merkmalswissen wird gelernt. In Experiment 1 durchliefen alle Versuchspersonen vor der ersten und vor der dritten Lerneinheit einen allgemeinen Wissenstest. Zu allen 15 Vogelarten wurde dabei das Merkmalswissen über vier komplexe, mehrere Einzelmerkmale enthaltende Fragen in Form eines Multiple-Choice-Tests erhoben. Maximal konnten in dem Test 60 Punkte erzielt werden. Die Versuchspersonen der beiden Experimentalgruppen verbesserten sich von Test 1 zu Test 2 von durchschnittlich 32 auf 45.25 Punkte. Diese Verbesserung ist im t -Test auf dem 0.1%-Niveau signifikant ($t(23) = -11.19, p < 0.001$). Auch die Kontrollpersonen zeigen eine Zunahme der Punkte im Fragebogen (28 auf 30.78 Punkte), im Vergleich zu der Lernergruppe ist dieser Effekt jedoch gering und nicht signifikant ($t(8) = -2.04, p = 0.076$).

In Experiment 2 wurde das Merkmalswissen der Versuchspersonen zu Beginn jeder der vier Sitzungen mit einem Zuordnungs-Fragebogen geprüft. Dabei galt ein Datum als gewusst, wenn die Angabe innerhalb einer 20%-Marke des tatsächlichen Wertes gemacht wurde. In der ersten Sitzung erzielten die Personen der Lerngruppe durchschnittlich 8 Punkte (von 45 möglichen), die der Kontrollgruppe durchschnittlich 9 Punkte. Während die Kontrollgruppe auf diesem Level verblieb (im zweiten Fragebogen sank der Durchschnittswert sogar vorübergehend auf 7 Punkte), verbesserten sich die Personen der Experimentalgruppe fast linear über 13 und 19 auf 22 Punkte. Diese Verbesserung verteilt sich gleichmäßig auf die drei gelernten

Merkmale und zeigte in einer zweifaktoriellen Varianzanalyse (ANOVA) mit den zwei Inter-subjektfaktoren Merkmal und Zeit signifikante Effekte des Merkmals, $F(1.78, 42.82) = 10.14$, $MSE = 86.30$, $p < .001$, und der Zeit, $F(2.0, 47.98) = 53.57$, $MSE = 457.44$, $p < .001$. Trotz dieser Verbesserung beherrschten die Versuchspersonen vor Beginn des vierten Lerndurchgangs aber erst ca. 50% des möglichen Gesamtwissens.

Die erzielten Punktzahlen zeigen deutlich an, dass die Personen der Experimentalgruppen tatsächlich an Merkmalswissen hinzugewonnen hatten, nicht aber die Kontrollpersonen. Damit ist die erste der zwei Grundvoraussetzungen für die Messung von Assimilations- und Akkomodationsleistungen in den beiden Experimenten gegeben.

Qualitative Verbesserung der Wissenskarten. Die zweite Voraussetzung besteht darin, dass sich der Zugewinn an Merkmalswissen auch auf die Wissenskarten auswirkt. Diese wird durch eine Analyse aller AverageLosses zur Zielstruktur überprüft: Nach jeder der Lerneinheiten nähern sich die individuellen Wissenskarten unmittelbar der Zielstruktur an, so dass sie am Ende des Experiments 1 bei einem AverageLoss von durchschnittlich 0.47 und in Experiment 2 bei einem AverageLoss von 0.30 enden (Abbildung 25). Diese Verbesserung ist im t -Test in Experiment 1 auf dem 1%- und im Experiment 2 auf dem 0,1%-Niveau signifikant (Experiment 1: $t(23) = 3.22$, $p < 0.01$; Experiment 2: $t(24) = 4.15$, $p < 0.001$). Interessant ist auch, dass ein Teil der Verbesserung in den Wissenskarten in der Woche zwischen den Sitzungen jeweils wieder verloren geht. Teilweise war der im AverageLoss sichtbare Lerneffekt bereits so gering, dass davon zu Beginn der nächsten Sitzung nichts mehr übrig ist. Zumeist überwiegt aber der Lerneffekt deutlich den Vergessenseffekt und fällt vor allem nach dem ersten Lerndurchgang in beiden Experimenten am größten aus. Das Muster von Verbesserung der Wissenskarten durch Merkmalslernen und Verschlechterung durch eine Woche Pause wird im Binomialtest auf dem 5%- (Experiment 1) und dem 1%-Niveau (Experiment 2) signifikant.

Insgesamt zeigt sich ein Niveauunterschied der Wissenskarten zwischen den beiden Experimenten: Das zum Thema Ornithologie erweist sich als wesentlich anspruchsvoller, und die AverageLosses liegen konsequent um ca. 0.15 höher als im klarer strukturierten Länderexperiment. Auch gelingt es den Versuchspersonen der Kontrollgruppe im Experiment nicht, eine qualitativ anständige Karte zu produzieren. Zwar sind die gemessenen AverageLosses zufallskritisch gegen das Übereinanderlegen strukturfremder Karten abgesichert (der kritische Wert würde auf dem 1%-Niveau bei ca. 0.95 liegen), doch ist die Abweichung eines Wertes um die 0.70 doch beträchtlich. Insgesamt bildeten die Versuchspersonen der Kontrollgruppen in beiden Experimenten durchgängig schlechtere Karten und verbesserten sich zu keinem Zeitpunkt signifikant. Das mehrfache Abgeben von Ähnlichkeitsurteilen allein ergibt also keine Verbesserung in den Wissenskarten.

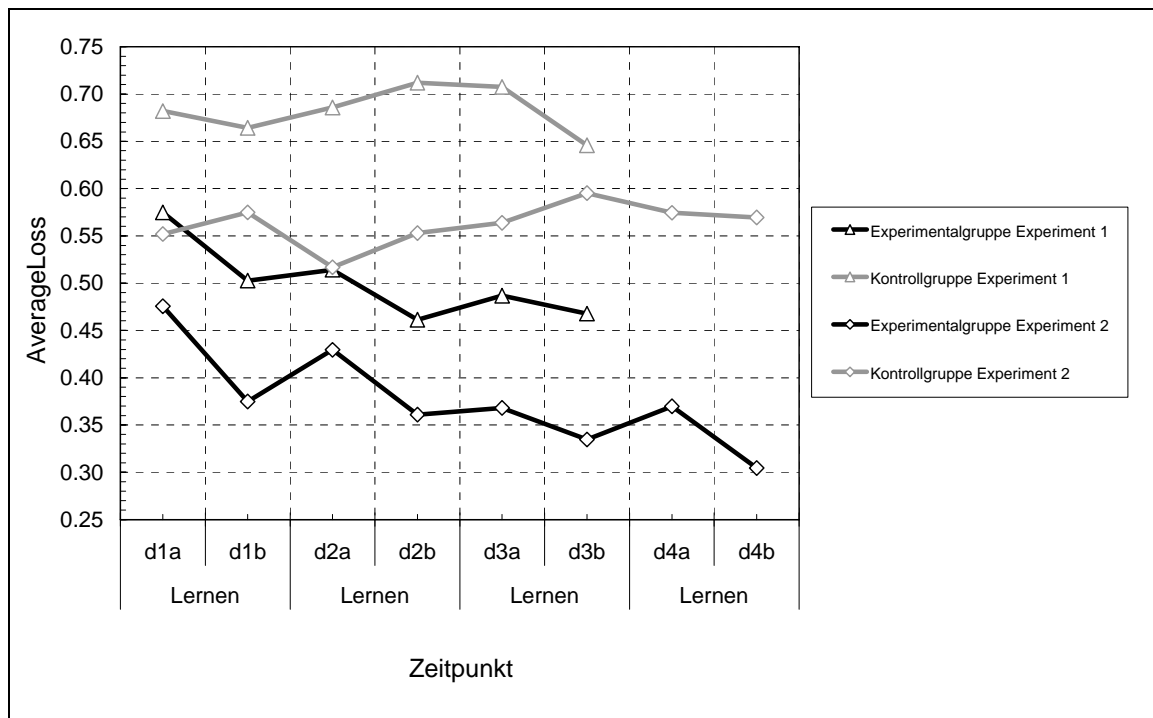


Abbildung 25. Gemittelte Abweichungen aller individuellen Wissenskarten von der Zielstruktur aus Experiment 1 und 2.

Übungseffekt bei wiederholter Abgabe von Ähnlichkeitsurteilen. Trotzdem machen die Kontrollpersonen nicht einfach „irgendetwas“. Dass sie sich bei den Ähnlichkeitsurteilen redlich Mühe geben, zeigt der jeweils erzielte Übungseffekt. Mit Ausnahme der vierten Lerneinheit in Experiment 2 verbessern sich die Versuchspersonen innerhalb einer Sitzung jeweils in der Konsistenz ihrer Urteile (Abbildung 26). Während der einwöchigen Pause geht dann (ebenfalls mit Ausnahme der letzten Sitzung von Experiment 2) der Konsistenzgewinn wie prognostiziert jeweils wieder zu einem guten Teil verloren. Das resultierende Muster entspricht – mit Ausnahme des in Experiment 2 zu niedrigen Mittelwertes für d4a – exakt den Vorhersagen. Zufallskritisch ist diese Übereinstimmung im Binomialtest auf dem 1%-Niveau als signifikant abgesichert (11 von 13 Vorhersagen korrekt bei einer Zufallserwartung von 50% pro Vorhersage). Auch hier erweist sich das Ornithologie-Experiment als das deutlich schwierigere der beiden. Nicht nur dass die Karten weiter vom Normmodell weg liegen, sie sind auch in sich um eine ganze Größenordnung inkonsistenter.

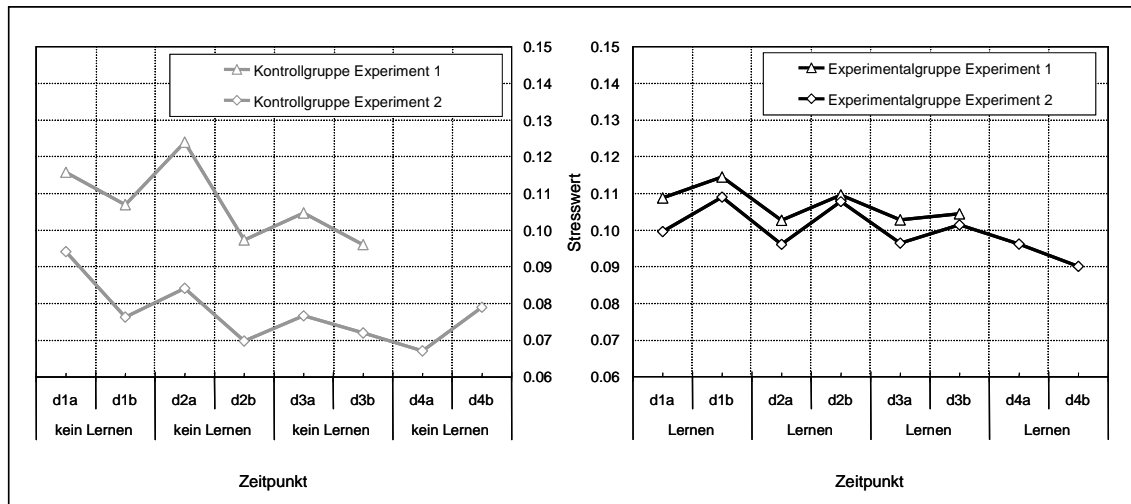


Abbildung 26. Stresswertverlauf der Kontrollgruppen (links) und der Experimentalgruppen (rechts) in Experiment 1 und 2.

Erhöhung der Inkonsistenzen durch Assimilationsprozesse. Für die Versuchspersonen der Experimentalgruppen haben wir genau gegenteilige Hypothesen hinsichtlich des Stresswertverlaufes aufgestellt. Ihre Karten sollten unmittelbar nach dem Lernen jeweils einen erhöhten Stresswert aufweisen. Das ist (wieder mit Ausnahme der vierten Sitzung (d4a, d4b) in Experiment 2) auch durchgängig der Fall (Abbildung 26).

Interessant an dem Vergleich der Kurven für die beiden Experimente ist, dass sie viel enger zusammen liegen als bei der Kontrollgruppe. Offensichtlich bietet das Lernmaterial aus Experiment 2 den Versuchspersonen die Chance, eine für sie konsistente Vorstellung über das Wissensgebiet aufzustellen (im Gegensatz zu den Kontrollpersonen, die ja nur auf ihr unzureichendes Vorwissen zurückgreifen können). Im Gegensatz dazu haben die Kontrollpersonen anscheinend in Experiment 2 einen Konsistenzvorteil. Sie können die Länder so strukturieren, wie sie das aufgrund ihres Vorwissens möchten, und müssen sich nicht mit der Aufgabe herumschlagen, das Lernmaterial zu integrieren. In der Summe ergibt sich dann der große Unterschied bei der Kontrollgruppe, aber nur ein minimaler Unterschied bei den Lernern.

Verringerung der Inkonsistenzen in den Lernpausen. Ebenfalls aus Abbildung 26 zu entnehmen ist die jeweilige Verbesserung der Konsistenz der Karten in der Woche zwischen den Sitzungen. Dies deckt sich mit der Annahme eine Gesamtintegration des hinzugewonnenen Wissens durch akkomodative Prozesse. Wie bereits bei den Personen der Kontrollgruppe, ist auch die Übereinstimmung des Datenverlaufs in den Experimentalgruppen (Abbildung 26) im Binomialtest auf dem 1%-Niveau abgesichert (12 von 13 Vorhersagen korrekt bei einer Zufallserwartung von 50% pro Vorhersage).

Bei der Betrachtung der einzelnen Vergleiche zwischen den Zeitpunkten, aufgeteilt nach Assimilation (Vergleich der Stresswerte bevor und nach dem Lernen) und Akkomodation (Vergleich der Stresswerte direkt nach dem Lernen und nach der Lernpause) zeigt sich dasselbe Bild (Tabelle 1). Bei der (den) Experimentalgruppe(n) wird bis auf eine Ausnahme jeder Einzelvergleich signifikant, wenngleich sich die Effektgrößen auf dem Niveau eines kleinen

($d = 0.2$) bis mittleren Effektes ($d = 0.5$). Bei der Kontrollgruppe wirkt sich die Übung positiv auf die interne Konsistenz der Urteile aus.

Tabelle 1. Überblick über Assimilations- und Akkomodationseffekte (Experimentalgruppe) und Übungs- und Vergessenseffekte (Kontrollgruppe) im Stresswert. Signifikanz der Mittelwertsvergleiche ($\alpha = 0.05$) = p(t); Effektstärke = d.

Stichprobe	Effekt	p(t)	d
Experiment 1 (Experimentalgruppe)	Assim	n.s.	0.14
	Akkom	s.	0.27
Experiment 2 (Experimentalgruppe)	Assim	s.	0.25
	Akkom	s.	0.35
Experiment 1 & 2 (Experimentalgruppe)	Assim	s.	0.20
	Akkom	s.	0.32
Experiment 1 (Kontrollgruppe)	Übung	s.	0.42
	Vergessen	n.s.	0.34
Experiment 2 (Kontrollgruppe)	Übung	s.	0.28
	Vergessen	n.s.	0.15
Experiment 1 & 2 (Kontrollgruppe)	Übung	s.	0.31
	Vergessen	s.	0.20

Schätzung der Absolutgröße des Assimilationseffektes und des Akkomodationseffektes. Die beiden in der Mittelwertstendenz aufzufindenden Effekte (die längst nicht für die Stresswerte jeder individuellen Versuchsperson gelten!) geben den prognostizierten Hinweis auf Assimilation und Akkomodation. Einleitend hatten wir festgestellt, dass die Erwartungslinie der Stresswerte in diesem Experiment allerdings nicht das Beibehalten des Stresswertes ist, sondern der Übungseffekt bzw. seine Auflösung bei Personen, die nur Ähnlichkeitsurteile abgeben, aber nicht lernen. Aus diesem Grund präsentiert Abbildung 27 eine Schätzung des „Netto-Effektes“, indem jeweils die zu überwindende gegenläufige Tendenz hinzuaddiert wird. Dabei ist allerdings in der Interpretation die gleiche Vorsicht angeraten wie beispielsweise bei hirnpfysiologischen Messungen (wo ja ebenfalls eine entsprechende Differenzbildung zum Aufdecken von Effekten betrieben wird), weil Stresswerte zwar im Prinzip intervallskaliert sind, aber nicht mit Sicherheit eine vollständig lineare Verlaufsfunktion über eine beliebige Anzahl erzeugter Stresswerte haben (es gibt unseres Wissens dazu keine systematischen Berechnungen oder Simulationen). Trotzdem kann man sich den Differenzeffekt ja einmal anschauen, denn er ergibt doch beachtliche Werte (mit Ausnahme für die vierte Sitzung in Experiment 2, in der die Versuchspersonen beider Gruppen, wie wir gesehen haben, systematisch vom ansonsten durchgängigen Muster abwichen).

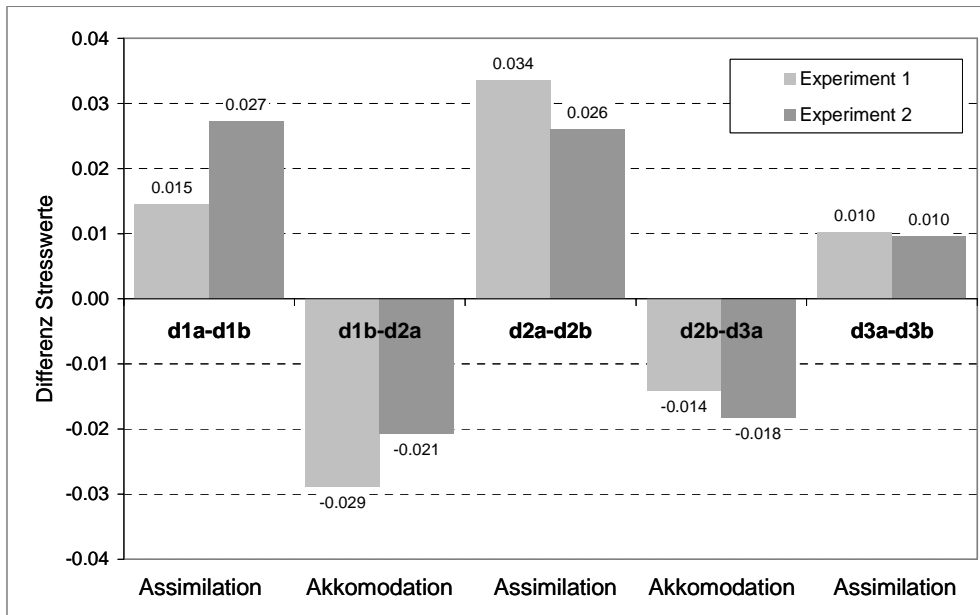


Abbildung 27. „Netto-Effekte“ der Assimilation und Akkomodation für Experiment 1 und 2 über die Zeit (Für Experiment 2 fehlt der letzte Akkomodations- (d3b-d4a) und der letzte Assimilationseffekt (d4a-d4b) aufgrund der fehlenden Berechnungsberechtigung; nähere Erläuterungen im Text). Absolute Assimilations- und Akkomodationseffekte wurden von Übungs- und Vergessenseffekten bereinigt.

3.4. Diskussion

Die Resultate der beiden Experimente zeigen einen mit den Vorhersagen erstaunlich gut übereinstimmenden Verlauf, und zwar sowohl in den Lernergruppen als auch in den Kontrollgruppen. Das spricht für einen systematischen Einfluss der Lerneinheiten auf die innere Konsistenz der Ähnlichkeitsurteile, und zwar ganz in der Richtung, wie dies durch die Prozesse von Assimilation und Akkomodation theoretisch beschrieben wird (bei gleichzeitig feststellbaren Lern- und Vergessenseffekten im Vergleich zu einer Wissensnorm!). Der eingesetzte robuste NMDS-Algorithmus scheint also die gewünschte „Filterfunktion“ zu erbringen, denn es bleibt jeweils das vorhergesagte unterschiedliche Ausmaß an Stress als nicht-skalierbares Rauschen der Daten in diesem „Filter“ hängen. Damit konnte dieses technische Maß zum ersten Mal für die psychologische Forschung auch inhaltlich sinnvoll eingesetzt werden.

Die einzige Unregelmäßigkeit im Muster des Stresswert-Verlaufs tauchte im letzten Lerndurchgang von Experiment 2 auf. Hier erwies sich der mittlere Stresswert nach dem Lernen als ungewöhnlich niedrig, was den erwarteten Effekt konterkarierte. Um die Stabilität respektive die Zufälligkeit des Stresswertverlaufs zu testen, haben wir die Lernergruppe von Experiment 2 in zwei Teilstichproben zerlegt. Beide Teilstichproben zeigen exakt einen parallelen Verlauf der Stresswerte über alle Zeitpunkte hinweg, inklusive eben dem letzten. Hier muss also etwas Systematisches im Aufbau des Experiments passiert sein, was nicht nur auf einige wenige Personen zurückzuführen sein kann (zumal ja auch die Resultate der Kontrollgruppe in diesem Durchgang genau gegen das sonst durchgängige Befundmuster laufen). Spekulieren könnte man über motivationale Probleme mit einem doch zum Schluss recht repetitiv werdenden Experiment. Aber dagegen spricht, dass die AverageLosses für die Experimentalgrup-

pe einen sehr ordentlichen Lerneffekt aufweisen. Es gibt eben Resultate, die sich einer vernünftigen Interpretation zu verschließen scheinen.

Für die Gesamtlage der Befunde kommt den Kontrollgruppen eine wichtige Funktion zu: *Ers-*
tens zeigen sie, dass der „normale“ Prozess der Konsistenz von Ähnlichkeitsurteilen gerade
umgekehrt ist: Die Wiederholung innerhalb einer Sitzung fördert die Konsistenz, die Woche
Pause nimmt Konsistenz weg. Damit ist *zweitens* aber ein additiver Effekt hinsichtlich des
Lernens gegeben: Die Stresswertverläufe der Experimentalgruppen wiegen nicht nur den
normalen Übungseffekt auf, sondern sie übertreffen ihn sogar in einem signifikant werdenden
Masse. Wenn man will, dann kann man die Beträge (Lernergruppe + Kontrollgruppe) also
ruhig addieren, um eine Schätzung zu haben, wie stark sich das Lernen bzw. die einwöchige
Pause danach auf Assimilation und Akkomodation auswirkt. Allerdings sind die vorgelegten
Befunde zunächst einmal eher als Resultate zu verstehen, die die Existenz eines systemati-
schen Stresswert-Effekts als solches belegen. Sie können keine endgültige quantitative Be-
zugsgröße liefern und sollen das auch gar nicht. Die Berechnung eines „Netto-Effekts“ ist
deswegen lediglich als ein Verfahrensvorschlag einzuordnen, um den kontrastierenden Ü-
bungseffekt korrekterweise mit ins Kalkül einzubeziehen. Hier wäre noch weitere systemati-
sche Forschung notwendig, welche die Größe von Assimilations- und Akkomodationseffekten
als unabhängige Variable definiert und dann den NMDS-Stresswert als abhängige Variable
quantitativ auslotet.

Die hier vorgelegten Befunde wurden auch lediglich mit einem statistisch eher schwachen
Test statistisch abgesichert: Die aus der Theorie abgeleiteten Prognosen hinsichtlich Vergrö-
ßerung und Verringerung hätten unter Zufallsbedingungen eine Eintretenswahrscheinlichkeit
von je 50% gehabt, so dass sich dieses Prognoseschema mittels Binomialtest auf dem 1%-
Niveau vom Zufall abheben ließ. Allerdings liegen die Mittelwerts-Unterschiede zwischen
zwei aufeinander folgenden Zeitpunkten liegen jeweils unter einer Standardabweichung der
beiden Verteilungen. Die Effektstärke ist also wie erwartet gering bis mittel; der Stresswert
wird halt durch die Summe der anderen Faktoren stärker beeinflusst als durch Akkomodation
und Assimilation.

Nun ist, wie in der Einleitung bereits ausgeführt, der Stresswert in allererster Linie ein inter-
nes technisches Maß für den Skalierungsprozess. Seine Höhe in einer NMDS-Lösung ist
durch die Restriktionen bei der iterativen ordinalen Darstellung der Proximitäten zwischen n
Objekten in einem niedrigdimensionalen Raum bedingt. Deswegen darf nicht damit gerechnet
werden, dass sich auf Einzelfallebene individueller Karten bereits vorhersagbare Unterschiede
zeigen. Insofern ist der Stresswert als psychologisches Maß vielleicht am besten mit Reakti-
onszeiten oder mit hirnpfysiologischen Massen zu vergleichen: Auch hier sind die einzelnen
Messungen von vielen Gegebenheiten abhängig, von denen die unabhängige Variable des
Experiments in der Regel nur einen geringen Teil ausmacht. Ein einzelnes Messdatum allein
sagt deswegen herzlich wenig aus und wird von der Forschung auch gar nicht beachtet. Erst
wenn man viele Messungen übereinander legt, zeigt sich in den Mittelwerten ein Effekt im
Sinne eines in seiner Größe stabilen und statistisch signifikanten Unterschiedes zu einer Ver-
gleichsmessung.

Obwohl in der Psychologie solche Form von Messqualität üblich ist, wäre an dieser Stelle eine Art „medizinischer Test“ natürlich wünschenswerter. Ein Test, der mit klar definierter Sensitivität und Spezifität für den Einzelfall eine Diagnose über Assimilation und Akkommodation zuließe, könnte nämlich Aussagen über den individuellen kognitiven Prozess einer lernenden Person machen. Die vorgelegten Ergebnisse müssen jedoch Zweifel daran aufkommen lassen, ob es möglich sein wird, aus der Konsistenz von Urteilen einen solchen Einzelfall-Test abzuleiten. Mit dem NMDS-Stresswert scheint uns das Limit auf Stichprobenebene bereits erreicht zu sein⁵.

Gleichwohl öffnet dieses Maß den Zugang zur Abschätzung von Assimilations- und Akkomodationsprozessen auf einer neuen Ebene, auf der des objektbezogenen Faktenwissens nämlich. Damit wird die bisherige Forschung, die sich hauptsächlich auf a) die Entwicklung mathematisch-naturwissenschaftlichen Wissens (für einen Überblick bzw. für instruktive Beispiele s. Inhelder, De Caprona, & Cornu-Wells, 1987; Inhelder, Sinclair, & Bovet, 1974; Sodian, 1995; Venville & Treagust, 1998; Vosniadou, 1991), auf b) die Modellierung und Veränderung propositionalen Sprachwissens (Aebli, 1976; Steiner, 1996) oder auf c) die Beschreibung der Genese sozialer Kognition (Eckensberger & Silbereisen, 1980; Freitag, 1984; Furth, 1980;), Interaktionskompetenz (Bardis, 1978, 1979; Edelstein & Habermas, 1984) und Integration (Chiswick, 2002; Demorgon & Molz, 1996; Ghuman, 1994; Kao & Tienda, 1995; Mayades & Elliott, 1992; Straub, 1999; Wang, 1991; Weinberg, 2001) bezog, um ein viertes Feld erweitert, auf das sich das Konzept Piagets ebenfalls sehr gut anwenden lässt.

Inhaltlich in jedem Fall interessant sind die Interpretationsperspektiven der hiermit bereits vorgelegten Resultate hinsichtlich der Wissensintegration beim Lernen von Sachwissen. In Verbindung mit der qualitativen Verbesserung der individuellen Wissenskarte zeigt der Assimilationseffekt direkt nach dem Lernen, dass die Personen die hinzugewonnenen Fakten zwar gewinnbringend einsetzen können, dass dieser Einsatz aber noch nicht in sich stimmig ist. Nachdem sie sich dann eine Woche nicht bewusst mit dieser Materie befasst haben, ist ein Teil des erworbenen Wissens wieder vergessen (qualitativ verschlechtern sich die Wissenskarten im Vergleich zum Normmodell wieder). Aber das, was die Personen behalten haben, ist jetzt als Gedächtnisspur offenbar so gut mit dem zuvor bereits bestehenden Wissen verschmolzen, dass die Gesamtkonsistenz der abgegebenen Ähnlichkeitsurteile merklich ansteigt. Hier bietet sich auch von gedächtnispsychologischer Seite her ein lohnenswert erschei-

⁵ Eine Alternative könnte man darin sehen, auf Matrixebene die ordinale Transitivität zu überprüfen. Diese ist aber – abgesehen von Spezialfällen – prinzipiell gegeben. Und die Spezialfälle allein eröffnen kein ausreichendes Feld für eine numerische Auswertung. Natürlich muss eine Person, um konsistent zu sein, in ihren Urteilen auch transitiv sein: Wenn die Ähnlichkeit zwischen A und B größer ist als zwischen C und D und die Ähnlichkeit zwischen C und D größer ist als zwischen E und F, dann muss auch die Ähnlichkeit zwischen A und B größer sein als zwischen E und F. Das ist aber immer der Fall, weil die Vp ja paarweise Urteile abgibt und nicht Paare miteinander vergleicht. Für einen Transitivitätsvergleich müsste eine Vp also immer Ähnlichkeitsurteile zwischen zwei Paaren von Objekten abgeben. Aber das sprengt kombinatorisch sofort die Grenze des praktisch Machbaren, weil bei 15 Objekten bereits mehr als 5000 Paare von Paaren möglich sind. Kaufmann (1999) und Läge (2001) legen einen Kurztest zur Abschätzung von Transitivität bei Paaren von Paaren vor, der allerdings für die feinen Unterschiede, so wie sie zwischen assimilativen und akkomodativen Prozessen zu erwarten sind, viel zu grob misst.

nender Ansatzpunkt, die Integrationsprozesse des Faktenwissens noch systematischer unter die Lupe zu nehmen als das in den zwei in dieser Arbeit präsentierten Experimenten möglich war.

Ein Festhalten an den Ähnlichkeitsurteilen als Instrument der Datenerhebung scheint dazu nicht der schlechteste Weg zu sein. Klauer (1989) hat das Bilden solcher Urteile als induktives Denken beschrieben und – im Duktus der von ihm herausgearbeiteten Transfereffekte – das Feststellen von Gleichheit und Verschiedenheit als Ausgangspunkt für das Entdecken von Regelmäßigkeiten benannt. Für den experimentellen Bereich ist dazu dieser Konnekt zu den Transferleistungen nützlich, aber nicht einmal unbedingt notwendig. Es reicht aus, mit Klauer festzustellen, dass bei Ähnlichkeitsurteilen „... nicht Objekte als Ganze und global, sondern Merkmale von Objekten beziehungsweise Relationen zwischen Objekten miteinander verglichen werden.“ (S. 187). Mit einem gesunden Schuss Optimismus könnte man demnach das Ähnlichkeitsurteil als eine Art „Königsweg zur Relationalität des Merkmalswissens“ bezeichnen.

Die mittels solcher Ähnlichkeitsurteile vorgelegten Resultate zeigen denn auch auf, dass die relational strukturierenden Prozesse von Assimilation und Akkomodation – so wie dies von Piaget angenommen wird – zwei (zeitlich voneinander verschiedene) Teilprozesse einer Gesamtfunktion sind, nämlich der, ein neues Gleichgewicht zwischen alten und neuen Wissens-elementen zu bilden. Vor diesem Hintergrund muss darauf verwiesen werden, dass die differentialdiagnostische Begriffsverwendung „assimilativer“ oder „akkomodativer“ Lernstile von Personen (z.B. Lynch, Woelfl, Steele, & Hanssen, 1998) maximal metaphorisch gesehen werden sollte. Unter gedächtnispsychologischer Perspektive hängen beide sehr eng miteinander zusammen, weil sie sich unwillkürlich, aber zeitversetzt aus demselben Lernen ergeben. Insofern bestätigen die Befunde Piagets Theorie der Äquilibration auf einer Ebene, auf der bislang wenig nach Struktur gesucht wurde. Das lässt eine weitere Erforschung der kognitiven Prozesse beim Lernen von Faktenwissen als lohnend erscheinen.

3.5. Literatur

- Aebli, H. (1976). *Psychologische Didaktik. Didaktische Auswertungen der Psychologie von Jean Piaget*. Stuttgart: Klett.
- Ausubel, D. P. (2000). *The acquisition and retention of knowledge*. Dordrecht: Kluwer Academic Publishers.
- Bardis, P. D. (1978). Social interaction and social processes. *International Behavioural Scientist*, 10, 7-32.
- Bardis, P. D. (1979). Social interaction and social processes. *Social Science*, 54 (3), 147-167.
- Borg, I. (1978). PAL: Point-wise alienation coefficients in multidimensional scaling. *Journal of Marketing Research*, 15, 478-479.
- Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling – Theory and applications*. New York: Springer.
- Chi, M. T. H., & Koeske, R. D. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology*, 19, 29-39.
- Chiswick, B. R., Lee, Y. L., & Miller, P. W. (2002). *Longitudinal analysis of immigrant occupational mobility: a test of the immigrant assimilation hypothesis*. Bonn: IZA.

- Demorgon, J., & Molz, M. (1996). Bedingungen und Auswirkungen der Analyse von Kultur(en) und interkulturellen Interaktionen. In A. Thomas (Hrsg.), *Psychologie interkulturellen Handelns* (S. 43-86). Göttingen: Hogrefe.
- Eckensberger, L. H., & Silbereisen, R. K. (Hrsg.). (1980). *Entwicklung sozialer Kognitionen: Modelle, Theorien, Methoden, Anwendung*. Stuttgart: Klett-Cotta.
- Edelstein, W., & Habermas, J. (Hrsg.). (1984). *Soziale Interaktion und soziales Verstehen. Beiträge zur Entwicklung der Interaktionskompetenz*. Frankfurt/M.: Suhrkamp.
- Freitag, B. (1984). *Der Aufbau kindlicher Bewusstseinsstrukturen im gesellschaftlichen Kontext*. München: Fink.
- Furth, H. G. (1980). *The world of grown-ups. Childrens conceptions of society* (Chapt. 9: Piagets theory of social relation). New York: Elsevier.
- Ghuman, P. A. S. (1994). Canadian or Indo-Canadian: A study of South Asian adolescents. *International Journal of Adolescence and Youth*, 4, 229-243.
- Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie*. München: Reinhardt.
- Inhelder, B., De Caprona, D., & Cornu-Wells, A. (Eds.). (1987). *Piaget Today*. Sussex: Erlbaum.
- Inhelder, B., Sinclair, H., & Bovet, M. (1974). *Apprentissage et structures de la connaissance*. Paris: Presses Universitaires de France.
- Kao, G., & Tienda, M. (1995). Optimism and achievement: the educational performance of immigrant youth. *Social Science Quarterly*, 76, 1-19.
- Klauer, K. J. (1989). Allgemeine oder bereichsspezifische Transfereffekte eines Denktrainings. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 21, 185-200.
- Läge, D. (2001). *Ähnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Läge, D., Daub, S., Bosia, L., Jäger, C. & Ryf, S. (2005). *Die Behandlung ausreißer-behafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag*. AKZ-Forschungsbericht Nr. 21. Zürich: Angewandte Kognitionspsychologie.
- Lynch, T. G., Woelfl, N. N., Steele, D. J., & Hanssen, C. S. (1998). Learning style influences student examination performance. *The American Journal of Surgery*, 176, 62-66.
- Marx, W., & Heij, A. (1989). *Subjektive Strukturen*. Göttingen: Hogrefe.
- Marx, W., & Läge, D. (1995). *Der ideologische Ring*. Göttingen: Hogrefe.
- Mathar, R. (1997). *Multidimensionale Skalierung*. Stuttgart: Teubner.
- Mayadas, N. S., & Elliott, D. (1992). Integration and xenophobia: an inherent conflict in international migration. *Journal of Multicultural Social Work*, 2, 47-62.
- Montada, L. (1995). Die geistige Entwicklung aus der Sicht Piagets. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (3. Aufl.) (S. 622-653). Weinheim: Psychologie Verlags Union.
- Novak, J. D. (2002). Meaningful learning: the essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, 86 (4), 548-571.
- Piaget, J. (1976). *Die Äquilibration der kognitiven Strukturen*. Stuttgart: Klett.
- Roth, W.-M. (2000). Artificial neural networks for modeling knowing and learning in science. *Journal of Research in Science Teaching*, 37 (1), 63-80.
- Sodian, B. (1995). Entwicklung bereichsspezifischen Wissens. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (3. Aufl.) (S. 622-653). Weinheim: Psychologie Verlags Union.
- Steiner, G. (1996). *Lernen – Zwanzig Szenarien aus dem Alltag* (2. vollst. überarb. Aufl.). Bern: Huber.
- Straub, J. (1999). *Verstehen, Kritik, Anerkennung. Das Eigene und das Fremde in der Erkenntnisbildung interpretativer Wissenschaften* (Essener Kulturwissenschaftliche Vorträge, Band 4). Göttingen: Wallstein.
- Strube, G. (1996). Wissen & Wissensrepräsentation. In G. Strube et al. (Hrsg.), *Wörterbuch der Kognitionswissenschaft* (S. 799-815). Stuttgart: Klett-Cotta.
- Venville, G. J., & Treagust, D. F. (1998). Exploring conceptual change in genetics using a multidimensional interpretic framework. *Journal of Research in Science Teaching*, 35 (9), 1031-1055.

- Vosniadou, S. (1991). Conceptual development in astronomy. In S. M. Glynn, R. H. Yeany, & B. K. Britton (Eds.), *The psychology of learning science* (pp. 149-178). Hillsdale, NJ: Erlbaum.
- Wang, L. L. (1991). Roots and changing identity of the Chinese in the United States. *Daedalus*, 120 (2), 181-206.
- Weinberg, N. (2001). Mass migration and labor market incorporation: Soviet immigrants in Israel. *Research in Social Stratification and Mobility*, 18, 249-274.

4.

**Der ObjLoss einer
Prokrustes-Transformation
als Maß für Wissen**

Zusammenfassung

Wenn zwei euklidische Wissenskarten gleicher Objekte (hier NMDS-Karten von Ähnlichkeitsurteilen) mittels Prokrustes-Transformation übereinander gelegt werden, dann ergibt sich für jedes korrespondierende Objekt in der Karte ein Abweichungswert (gemessen als ObjectLoss). Wird dieser Abweichungswert zwischen der Wissenskarte einer Person und einem Expertenmodell erhoben, dann lässt sich daraus für jedes einzelne Objekt der Karte abschätzen, wie gut diese Person dieses Objekt einzuordnen in der Lage ist. Der vorliegende Beitrag lotet Möglichkeiten und Grenzen dieser Diagnostikidee anhand einer Simulationsstudie und anhand eines Lernexperimentes aus. Von technischer Seite ist dabei der mögliche Anteil an Objekten kritisch, die nicht-gewusst sind und mit ihren möglicherweise äußerst inkonsistenten Urteilen die Stabilität der Wissenskarte belasten. Dieser Anteil darf laut Simulationsstudie bis zu 40% der gesamten Objektmenge betragen. Im Lernexperiment erreichte die Diagnosequalität allerdings erst dann eine befriedigende Größenordnung, wenn weniger als 20% der Objekte nicht gelernt worden waren.

Schlüsselwörter: Nonmetrische Multidimensionale Skalierung, Prokrustes-Transformation, Faktenwissen, Wissensdiagnose, Kognitive Karte, Kognitive Struktur, Intelligentes Tutorielles System, Adaptives Lernsystem

4.1. Einleitung

Die Entwicklung computergestützter Ausbildungseinheiten wird absehbar zu einer zunehmenden Individualisierung von Lernvorgängen führen: Musste im gewöhnlichen Klassenunterricht gleichzeitig auf den Kenntnisstand aller Lernenden Rücksicht genommen werden, so ermöglicht ein virtueller Unterricht eine stärkere Hinwendung auf das aktuelle Wissen des Einzelnen, und zwar nicht nur hinsichtlich des anzuschlagenden Lerntempos, sondern auch in Bezug auf Auswahl und Zusammenstellung der einzelnen Lektionen. In einem derartigen „adaptiven Unterricht“ liegt eine der großen Chancen virtueller Lernprogramme.

Doch ist ein solches Ziel nicht durch eine einseitige Fokussierung auf das Erstellen multimedialen Unterrichtsmaterials zu erreichen. Vielmehr wird (gewissermaßen als zweites Standbein) eine verlässliche und effiziente Statusdiagnose des aktuellen Wissensstandes eines Lernenden benötigt. Leutner (1992b, 1993) hat bereits darauf hingewiesen, dass man sich bei einer solchen lernprozessbegleitenden Diagnose des Wissens einem Dilemma zu stellen hat: Auf der einen Seite soll eine möglichst umfassende und genaue Diagnose erstellt werden, um bei der Zusammenstellung der folgenden Lektionen so adaptiv wie möglich sein zu können. Auf der anderen Seite will man aber nicht zuviel Zeit und kognitive Ressourcen der Lernenden für die Diagnose beanspruchen, weil beides von der verbleibenden Lernzeit abzuziehen ist.

In den heutigen Entwürfen zum virtuellen Unterricht wird dieser Konflikt zumeist nicht als Optimierungsproblem begriffen, sondern zu der einen oder der anderen Seite hin aufgelöst: Entweder wird das Gelernte (in ungefähr derselben Form) nach jeder Lektion noch einmal

abgefragt, und die Zulassung zur nächsten Lektion geschieht nur bei einer entsprechenden Gesamtleistung im Test (das ist nicht sonderlich effizient). Oder es bleibt dem Lernenden überlassen, Lerntempo und Auswahl der Lektionen zu bestimmen (was äußerst effizient, aber schwerlich sonderlich genau sein dürfte, weil ein Lernender zwar eigene Wissenslücken, aber nicht Fehler im Wissen einzuschätzen in der Lage ist).

In dieser bestehenden Situation ist es zur Aufgabe der Wissenspsychologie geworden, Diagnostikverfahren zu generieren, die sowohl hinsichtlich der Genauigkeit als auch hinsichtlich der Effizienz einen für die Praxis brauchbaren Weg darstellen. Bereits gut ausgebaut ist dabei die Diagnostik von Vorgehensstrategien und Schlussfolgerungsregeln. Neben den Erfahrungen aus der Expertenforschung verfügt man mit Textaufgaben, wie sie aus dem Mathematik- und Physikunterricht bestens vertraut sind, über ein altbewährtes Diagnose-Instrument: Mit wenigen, geschickt ausgewählten Aufgaben kann das Repertoire einer Person an *Vorgehens-, Transformations- und Entscheidungsregeln* vollständig ausgelotet werden. Drei Beispiele aus jüngerer Zeit dafür bieten das mit großem Aufwand an der Universität Heidelberg erarbeitete und nun in Graz fortgeführte Verfahren zur Diagnostik von Schülerkompetenzen in der Wahrscheinlichkeitsrechnung (vgl. Albert & Lukas, 1999), die interdisziplinäre Zusammenarbeit der Instruktionspsychologie der Universität Erfurt mit der Mediziner Ausbildung in Sachen Diagnostiktraining im Fach Radiologie (Leutner et al., 1997; Brünken et al., 1997) und das an der School of Computer Studies in Leeds entwickelte INTUITION-System makro- und mikroökonomischer Handlungsweisen in der Wirtschaft (Siemer & Angelides, 1998). Alle drei Beispiele zeigen, wie im Grundzug regelbasiertes Wissen auf individueller Ebene erkannt und gefördert werden kann.

Anders sieht dies bei der Diagnostik von *merkmalsbezogenem Sachwissen* aus. Geht es um all die lästigen, aber für die Qualität von Expertenentscheidungen eben bitter notwendigen einzelnen Fakten, so müsste man nach und nach den gesamten Wissensbestand eines Faches in Tausenden von Fragen abtesten. Nur so wäre herauszufinden, was eine Person genau weiß und was nicht. Will ein Biologielehrer beispielsweise das zoologische Wissen eines Schülers erkunden, so muss er zu jeder Tierart alle Einzelmerkmale und hierarchischen Verknüpfungen abfragen, die dieses Tier kennzeichnet.

Das ist ein in der Praxis nicht durchführbares Unterfangen. Eine einfache Stichprobenziehung aus den möglichen Fragen, die statt dessen in aller Regel für die Sachwissensdiagnostik durchgeführt wird, kann zwar Information für ein Gesamturteil über den gesamthaften Wissensstand bieten (ausdrückbar z.B. in Form einer Schulnote), dürfte aber kaum hinreichend sein, um den Wissensstand zu allen einzelnen Objekten eines Sachgebietes detailliert zu beschreiben.

Ein zweites, in unseren Augen genauso zentrales Problem bei der Diagnose von Sachwissen besteht darin, dass die Elemente dieses Wissens nicht einfach im Gedächtnis unverbunden nebeneinander stehen. Vielmehr bilden sie – gleich auf welchem Wissensstand – eine Art System, d.h. die Elemente konstituieren unwillkürlich einen Satz von Relationen. Jüngere Entwicklungen in der Wissenspsychologie haben – z.T. unter Rückgriff auf Konzeptionen eines Semantischen Netzwerkes (wie das von Quillian, 1968) – diese relationale Facette von

Sachwissen aufgegriffen und bilden sie mit so genannten Mapping-Techniken ab. Damit tragen sie dem Umstand Rechnung, dass sich Wissensstrukturen durch Differenziertheit, Vollständigkeit und Komplexität auszeichnen, also weit über eine einfache Anhäufung unverbundener Fakten hinausgehen (vgl. Mandl & Fischer 2000). Dabei bedienen sich die Mapping-Techniken vorwiegend graphischer Verfahren in der Datenerhebung. Exemplarisch zu nennen sind die zahlreichen Arbeiten mit der Heidelberger Strukturlegetechnik (basierend auf Scheele & Groeben, 1984) und die verwandten Verfahren der graphentheoretischen Darstellung von Wissensstrukturen (Bonato, 1990), die Netzwerk-Elaborierungs-Technik (Eckert, 1998), das Concept Mapping Software Tool (COMASOTO) zur Diagnose strukturellen Wissens (Weber & Schumann, 2000), die Wissensvermittlung, tutoriell unterstützte Wissensanwendung und Wissensdiagnose mit Begriffsnetzen nach Plötzner, Beller & Härder (2000) oder die Knowledge-Tracking-Methode von Janetzko & Strube (2000).

Mandl & Fischer (2000) resümieren als Stand der Forschung, dass nach anfänglichen Problemen bei der Wissensmodellierung (hoher Idiosynkrasie-Anteil, fragliche Validität und mangelnde Re-Test-Reliabilität) inzwischen theoretische Grundlagen und auch konkret eingesetzte Diagnoseverfahren und Benutzerinterfaces vorliegen. Sie verweisen in ihrem historischen Abriss auch auf die psychometrisch-strukturelle Perspektive, die „... mit verschiedenen Formen der Ähnlichkeitsskalierungen und ihrer graphischen Darstellung (z.B. Multidimensionale Skalierung, Clusteranalysen) ... [den Versuch unternehmen] ... psychische Phänomene und deren Veränderung mathematisch zu quantifizieren.“ (S. 4). Über die vorhandenen graphischen Mapping-Techniken hinaus sprechen die Autoren auch explizit die Möglichkeit an, auf der Basis nicht-graphischer Daten (z.B. multiple Paarvergleiche) über komplexe Algorithmen auf Kognitive Wissensstrukturen zu schließen.

Im Folgenden wird ein Weg zur Diagnose merkmalsbezogenen Sachwissens vorgestellt, der genau einer solchen Grundidee folgt: Über numerische relationale Urteile soll für einen Objektbereich festgestellt werden, welche der Objekte als unzureichend gewusst angenommen werden müssen. Grundlegendes Messinstrument dafür sind paarweise Ähnlichkeitsurteile.

Bereits mit wenigen Informationen sind wir Menschen in der Lage, ein ungefähres Urteil über die Ähnlichkeit zweier Dinge oder Personen abzugeben (vgl. Klauer, 1989). Dabei vertrauen wir offenbar auf unser „gutes Auge“, mit dem wir wichtige von unwichtigen Merkmalen unterscheiden und die verbleibenden wichtigen Merkmale in eine wohlproportionierte Relation zueinander bringen. Laien sind genauso in der Lage wie Experten, die Ähnlichkeit zwischen zwei Dingen abzuschätzen – auch wenn das Urteil natürlich beim Experten qualitativ besser ausfallen sollte.

Mit der Kennzeichnung des Ähnlichkeitsurteils als Pauschalwert für einen Merkmalsvergleich stehen wir auf dem Boden gedächtnispsychologischer Modelle, wie sie etwa in der Tradition von Smith, Shoben & Rips (1974) zur Erklärung von Prozessen der Objekt-Kategorie-Identifikation eingebracht wurden. Auch wir begreifen ein Ähnlichkeitsurteil als das Resultat einer mehr oder minder automatisiert ablaufenden Berücksichtigung verschiedener Merkmale und deren Ausprägungsunterschiede bei den zu beurteilenden Objekten. Anzahl und Gewich-

tung herangezogener Merkmale können dabei extrem variieren, je nach Kenntnisstand und kognitivem Stil einer Person.

Nach unserem Dafürhalten teilen Ähnlichkeitsurteile mit den bestehenden Mapping-Techniken gegenüber einer reinen Abfrageliste von Sachfragen den Vorteil, dass sie relationale Gefüge, Integration und Elaboration (Gewichtungen, kategoriale Strukturierung usw.) der Wissensstruktur mit messen. Zusätzlich sind sie unabhängig von Frageformaten und relationalen Vorschriften. Allerdings liegt die resultierende relationale Information bei Ähnlichkeitsurteilen genauso wenig unmittelbar auf der Hand wie das zu ermittelnde Merkmalswissen über die einzelnen Objekte (welches mit Mapping-Techniken nur umständlich zu erheben ist). Es bedarf vielmehr komplexer Auswertealgorithmen, um relationales Wissen und Merkmalswissen aufzuzeigen.

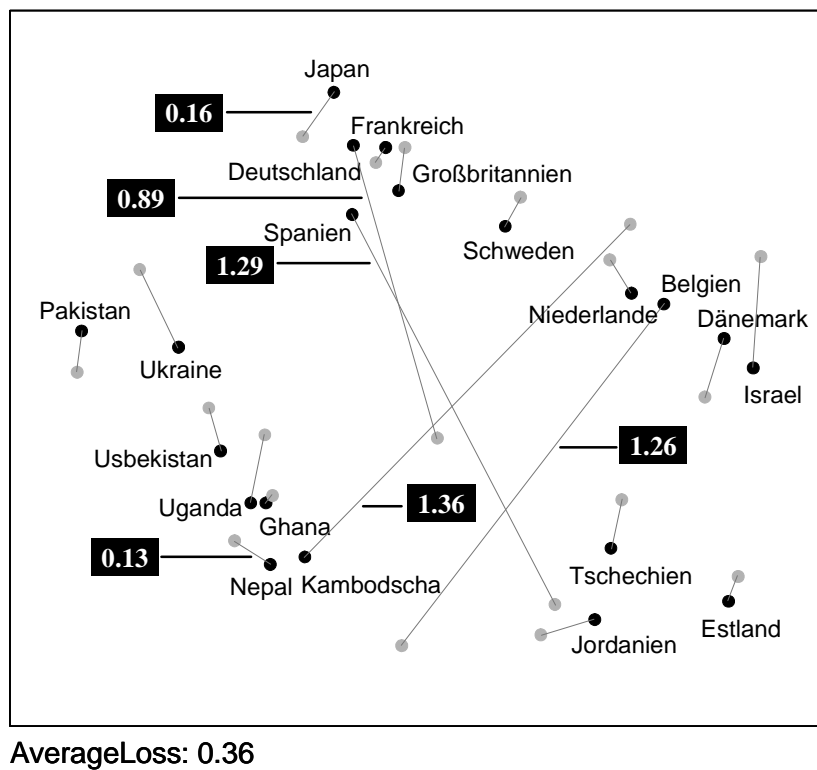


Abbildung 28. Ergebnis einer Prokrustes-Transformation zwischen der hypothetischen Wissenskarte einer Person (graue Punkte) und einer Normkarte (schwarze Punkte). Belgien, Deutschland, Kambodscha und Spanien sind mangelhaft platzierte Länder, die in einer Wissensdiagnose als "nicht gewusst" interpretiert werden. Die eingetragenen Werte sind die ObjectLosses. Sie bezeichnen die Abweichung zwischen je zwei korrespondierenden Punkten. Die Gesamtabweichung der beiden Karten voneinander ist durch den AverageLoss angegeben (hier: 0.36), der den Durchschnittswert aller ObjectLosses darstellt.

Durch Ähnlichkeitsurteile ermittelte Proximitäten zwischen Paaren von Objekten lassen sich mittels Nonmetrischer Multidimensionaler Skalierung (NMDS) als Distanzwerte verrechnen und zur Grundlage von „Wissenskarten“ machen (die Logik dieses Verfahrens ist z.B. bei Borg & Groenen, 1997 beschrieben). Auf diese Weise gewinnt man die relationale Information über das aktuelle Sachwissen. Solche individuellen „Landkarten“ können durch Prokrustes-Transformation mit einem Expertenmodell verglichen werden (Gower & Dijksterhuis,

2004). Bereits ohne großen mathematischen Aufwand springt bei einem solchen Vergleich ins Auge, welche Objekte ein Lerner gut kennt (d.h. korrekt verortet) und welche er falsch platziert (als Beispiel vgl. Abbildung 28). Doch natürlich lässt sich diese Abstandsinformation zwischen Ist und Soll auch numerisch ausdrücken (als „ObjectLoss“) und somit einer vollständig automatisierten Auswertung zugänglich machen (welche für eine Diagnose im Rahmen computergestützten Unterrichts ja wünschenswert wäre).

Ziel der vorliegenden Studie ist es, den gerade in knapper Form skizzierten Weg, um nicht-gewusste Objekte in einer Wissensstruktur automatisiert zu detektieren, auf seine Messgenauigkeit hin zu untersuchen. Die Möglichkeiten, aber auch die Grenzen eines solchen Umgangs mit der relationalen Seite der Information sollen dabei deutlich werden. Darüber hinaus besteht auch noch die Möglichkeit, aus der Position eines Objektes in der individuellen Wissensstruktur unmittelbar auf das dahinter stehende Sachwissen zu schließen. Dieser Weg kann mittels Kohonen-Netzen beschritten werden (Läge & Daub, 2005; Streule, Daub & Läge, 2005; vgl. auch Kapitel 6 in dieser Arbeit), stellt jedoch einen separaten Ansatz dar.

In der hier präsentierten Studie wird es ausschließlich darum gehen, gut gewusste von nicht-gewussten Objekten allein aufgrund ihrer Position in einer Kognitiven Karte zu unterscheiden. Als einziges zentrales Maß wird dazu der ObjectLoss herangezogen, der sich bei der Prokrustes-Transformation einer NMDS-Karte des individuellen Wissens mit einer NMDS-Karte des Expertenwissens ergibt. Für nicht-gewusste Objekte muss der ObjectLoss systematisch größer ausfallen als für gut gewusste Objekte, damit allein aus diesem numerischen Wert valide auf die Qualität des Wissens geschlossen werden kann.

Dabei sind zwei zentrale Schwierigkeiten zu überwinden. Die erste ist technischer Natur: Die eingesetzten statistischen Verfahren (NMDS und Prokrustes-Transformation) müssen in der Lage sein, die Karten vor allem aufgrund der gewussten Objekte aufzubauen bzw. übereinander zu legen und sich nicht durch deplatzierte Objekte „irritieren“ zu lassen. Diesen Aspekt bezeichnet man als Robustheit. Da normalerweise die Fehlermodelle von iterativen Approximationsalgorithmen (zu denen NMDS und Prokrustes-Transformation zählen) exklusiv auf Fehlerminimierung ausgelegt sind und nicht auf Robustheit, mussten erst entsprechende robuste Verfahren entwickelt werden (Läge, 2001; Läge, Daub, Bosia, Jäger, & Ryf, 2005).

Die zweite zentrale Schwierigkeit hängt inhaltlich mit dem Wissensstand einer Person zusammen: Gesetzt den Fall, eine Person würde über gar kein adäquates Wissen in einem Objektbereich verfügen, so wären die abgegebenen paarweisen Ähnlichkeitsurteile ohne jede innere Kohärenz (es fehlen ja die übergreifenden Merkmalsinformationen, um konsistent urteilen zu können). Die aus solchen Ähnlichkeitsurteilen resultierende Karte wäre chaotisch, und entsprechend würde die Prokrustes-Transformation mit dem Expertenmodell eine zufällige Ordnung der ObjectLosses ergeben. Gleiches muss man auch erwarten, wenn eine Person nur einige wenige Objekte kennt: Statistische Robustheit kann hier noch nicht greifen, weil zu wenig „Signal“ (= Wissen) vorliegt, um das „Rauschen“ (= Nichtwissen) in den Daten als ein solches zu identifizieren. Ohne ein Minimum an Stabilität in der Karte liegen dann auch die gewussten Objekte eher zufällig einmal an der richtigen Position, aber nicht systematisch.

Erst ab einem bestimmten, a priori nicht festlegbaren Wissensstand wird das Verhältnis von Signal und Rauschen so sein, dass das vorgeschlagene Verfahren greifen kann.

Diese „kritische Masse“ im Wissensstand einer Person stellt die Grenze dar, ab der eine Diagnostik mit dem vorgeschlagenen Verfahren sinnvoll wird. Um diese Grenze ausloten zu können, haben wir uns bei der Evaluation zu einem doppelspurigen Vorgehen entschlossen:

Zunächst werden die technischen Möglichkeiten und Grenzen des Verfahrens in einer systematischen Simulationsstudie ausgetestet. Dabei wird – für unterschiedlich große Objektmengen und unterschiedliche Mächtigkeit des bereits vorhandenen Wissens – jeweils überprüft, wie weit der ObjectLoss aus der Prokrustes-Transformation in der Lage ist, als abhängige Variable die noch nicht gewussten Objekte als solche zu identifizieren. Die Ähnlichkeitsurteile werden dazu künstlich berechnet.

In einem zweiten Schritt soll dann auf experimentellem Weg geprüft werden, ob es auch möglich ist, aus den faktisch abgegebenen Ähnlichkeitsurteilen einer Person auf gewusste und auf nicht-gewusste Objekte zu schließen. Dazu lernen Personen sequentiell eine bestimmte Objektmenge und werden beständig über ihr Wissen getestet. Gleichzeitig ist das experimentelle Setting so aufgebaut, dass die „kritische Masse“ des Wissens ermittelt werden kann, die für eine hinlänglich gute Vorhersage notwendig ist. Anders als bei einer Simulationsstudie ist es bei einem derartig aufwändigen Lern- und Diagnostikexperiment allerdings nur möglich, das für einen exemplarischen Fall mit einer einzigen Objektmenge durchzuspielen.

Mit der getroffenen Unterscheidung zwischen gewussten und nicht-gewussten Objekten haben wir eine dritte Kategorie unterschlagen, nämlich die „falsch-gewussten“: Eine Person kann subjektiv davon überzeugt sein, ein Objekt zu kennen. Objektiv betrachtet erweist sich dieses Sachwissen aber als inhaltlich inadäquat. In einem solchen Fall wird eine Person zwischen diesem falsch gewussten Objekt und den anderen von ihr gewussten Objekten sehr konsistente Ähnlichkeitsurteile abgeben; sie kann sich ja auf ihr Merkmalswissen stützen. Entsprechend belasten falsch gewusste Objekte die Stabilität einer Wissenskarte genau so wenig wie korrekt gewusste – nur werden sie halt am falschen Ort platziert sein. Je nachdem, wie inadäquat das falsche Wissen ist, wird der korrespondierende Punkt in der Karte weiter oder weniger weit von der korrekten Position entfernt sein. Dies wird durch den ObjectLoss automatisch registriert, und wenn der Wert eine bestimmte kritische Schwelle übersteigt, dann wird das Objekt in die Liste des noch zu Lernenden aufgenommen. Solche Fälle sind für die praktische Funktionsfähigkeit des vorgeschlagenen Verfahrens also nicht kritisch. Simulation und Experiment konzentrieren sich daher auf nicht-gewusste Objekte, weil nur diese die Wissenskarten instabil werden lassen können.

4.2. *Simulation*

4.2.1. *Methode*

Durchführung. Die Simulationsstudie basiert auf NMDS-Wissenskarten, die aus paarweisen Unähnlichkeiten zwischen einer Menge von Objekten berechnet sind. Als Inhaltsgebiet wurde

dazu ein Ausschnitt aus dem Wissensbereich der Geographie verwendet. Relevante Merkmalsdimensionen der Objektkategorie „Länder“ waren das Bruttosozialprodukt pro Kopf, die Einwohnerzahl und die Fläche. Alle Daten stammen aus Fischers Weltalmanach (2002) und beziehen sich auf das Jahr 2000. Es wurde darauf geachtet, durch die Objektauswahl auf allen drei Merkmalsdimensionen jeweils eine hinreichend große Varianz zu erzeugen, so dass jedes Merkmal in jeder simulierten Wissenskarte auch zum Tragen kam.

Diese drei intervallskalierten Merkmale dienten als Basis für die Berechnung von korrekten Normkarten. Dazu wurden in einem ersten Schritt Proximitätsmatrizen berechnet. Die Konstruktion jeder so genannten 2-Way 1-Mode-Matrix paarweiser Unähnlichkeiten zwischen den Länderpaaren geschah dabei durch additive Verrechnung der Größenverhältnisse auf den drei Merkmalsskalen:

$$U\ddot{A}_{x,y} = \frac{|(x_i - y_i)/(x_i + y_i)| + |(x_j - y_j)/(x_j + y_j)| + |(x_k - y_k)/(x_k + y_k)|}{3}$$

Wobei: $U\ddot{A}_{x,y}$: Unähnlichkeitsquotient zwischen zwei Objekten
 x, y : Länderindex
 i : Bruttosozialprodukt pro Kopf in US\$
 j : Einwohnerzahl
 k : Fläche in km²

Die so berechneten Unähnlichkeitsquotienten eines Länderpaares liegen numerisch zwischen 0 und 1, wobei der Wert 0 für perfekte Übereinstimmung auf allen drei Merkmalsvektoren stehen würde, während ein Unähnlichkeitswert von 1 ein hypothetischer Limes ist, der nur bei unendlich großen Werten überhaupt erreicht werden könnte. Diese Werte wurden invertiert und linear zwischen 1 und 9 skaliert (Die auf diese Weise erzeugte Skala entspricht den weiter unten im Experiment berichteten Ähnlichkeitsurteilen.).

Die durch dieses Vorgehen generierten Matrizen bildeten die Rohdaten für die Berechnung der Normkarten und für die Simulation von Nicht-Wissen. Die Normkarten wurden durch zweidimensionale NMDS dieser Matrizen mittels des robusten Algorithmus RobuScal erstellt.

Inkorrektes Wissen wurde durch zufällige Veränderung aller Matrixwerte des betreffenden Objektes realisiert. Damit wird, wie in der Einleitung beschrieben, „Unwissen“ simuliert, nicht aber ein explizit falsches Wissen. Weiß eine Person nämlich nichts über ein Objekt, so muss sie entweder raten oder sich auf periphere, im Sinne der Normkarte jeweils irrelevante Details stützen. Das sollte zu inkorrekten und vor allem auch zu inkonsistenten Urteilen führen, so wie dies durch die zufällige Veränderung der Matrixwerte am treffendsten simuliert wird.

In jedem Simulationsdurchgang wurden die Matrixwerte sequentiell verändert, zunächst für ein einzelnes Objekt, dann für ein zweites, für ein drittes etc. bis dass die gesamte Matrix verfremdet war. Auf diese Weise ließ sich durch NMDS und Prokrustes-Transformation dann exakt aussagen, bei wie viel nicht gewussten Objekten die Vorhersage über den ObjectLoss funktionierte oder eben nicht mehr funktionierte (Abbildung 29).

Die auf der Basis jeder veränderten Matrix berechnete zweidimensionale NMDS-Karte ließ sich mittels Prokrustes-Transformation auf die Karte der korrekten Matrix legen. Die resultierenden ObjectLosses ergaben dann für jedes Land die Schätzung, ob es sich um ein bereits gewusstes oder um ein nicht-gewusstes Land handelte (die kritische Grenze für die Vorhersage wurde jeweils bei n veränderten Ländern zwischen dem n -ten und $n+1$ sten ObjectLoss gezogen.)

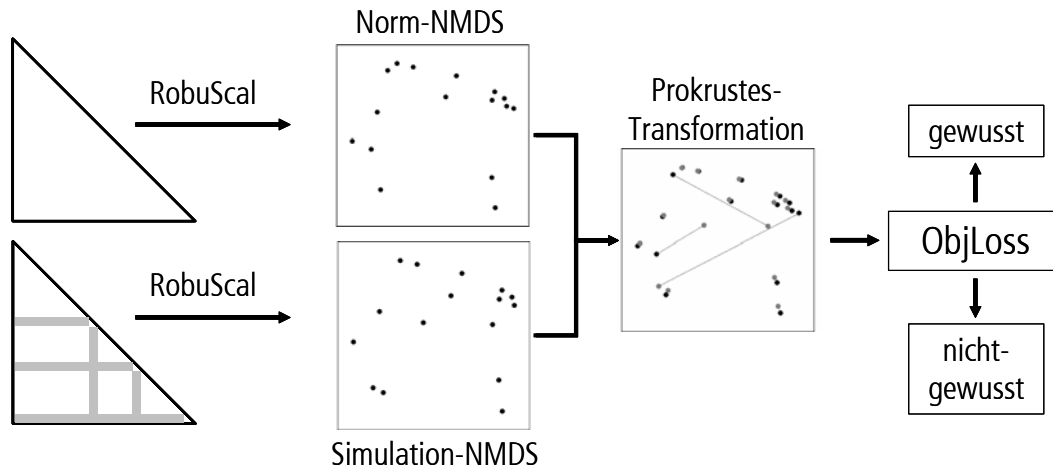


Abbildung 29. Schematische Darstellung des Simulationsablaufs. Die Matrizen links unten ($n = 12, 14, 16, 20, 24, 36$) werden pro Objekt sequentiell mit Zufallswerten versehen bis alle Werte verändert sind. Jede dieser Matrizen wird mit RobuScal skaliert und anschließend mittels einer Prokrustes-Transformation auf die NMDS-Karte gelegt, die sich aus der unveränderten Matrix ergibt. Daraus resultiert der ObjLoss als numerische Abweichung je zweier korrespondierender Objekte (dargestellt als Verbindungslinie).

Simuliert wurden jeweils 100 Wissenskarten unterschiedlicher Größe, bestehend aus 12, 14, 16, 20, 24 und 36 Ländern. Die Untergrenze von 12 Objekten folgt unseren Erfahrungen mit NMDS-Karten: In Erweiterung der alten Empfehlung von Shepard (1974) und Young (1970), mindestens 7 respektive 8 Objekte für eine zweidimensionale NMDS zu verwenden, konnten wir feststellen, dass bei kleinen Matrizen bereits die Änderung eines einzigen Wertes die resultierende Karte maßgeblich beeinflussen kann. Ab 12 Elementen ist eine NMDS-Karte aber hinsichtlich dieses Kriteriums stabil. Läge et al. (2005) konnten allerdings für 12 Objekte noch keinen hinlänglichen Effekt der Robustheit ausmachen, wohl aber bei 16 Elementen. Deswegen wird in diese Simulationsstudie noch die Objektzahl 14 eingefügt, um den kritischen Grenzbereich besonders gründlich auszuleuchten.

Karten einer Größe von deutlich mehr als 36 Objekten entbehren für die Praxis einer ökonomischen Handhabbarkeit (36 Objekte bedingen bereits 630 Ähnlichkeitsurteile). Für große Objektmenen werden wir deswegen ein anderes Verfahren vorschlagen (Läge, Schlatter & Ryf, 2006).

Hypothesen. Das Ziel der Simulation besteht darin, die Trefferrate des ObjectLosses als Maß für Unwissen abzuschätzen. Dabei wird ein großer ObjectLoss als Indiz für Nicht-Wissen gewertet, ein kleiner ObjectLoss hingegen als Indiz für Wissen. Diese Zuordnung ist technisch nicht so trivial wie es auf den ersten Blick erscheinen mag, weil ja an jedem Urteil (und

damit auch an jedem veränderten Matrixwert) immer zwei Objekte beteiligt sind. Deswegen ist nicht garantiert, dass es gerade die ungewussten Objekte sind, welche in der Karte eine andere Position bekommen; auch diejenigen Objekte, welche eigentlich „gewusst“ sind, erhalten in der Matrix ja notwendigerweise einige veränderte Proximitätswerte und könnten sich in der nachfolgenden NMDS deswegen genau so gut verschieben.

Außerdem wird ab einem bestimmten Ausmaß an verschobenen Objekten gar keine gemeinsame Grundstruktur mehr in der Punktekonfiguration der beiden NMDS-Lösungen enthalten sein, so dass eine Prokrustes-Transformation diese auch nicht mehr herausfiltern kann. Ab diesem Punkte bricht also das Verfahren als Diagnoseinstrument vollständig zusammen.

Aus diesen beiden Erwägungen formulieren wir folgende zwei Erwartungen, auf die jedes Teilresultat der Simulation zu überprüfen ist:

Kriterium 1: Die in der Proximitätenmatrix veränderten (= „nicht-gewussten“) Objekte haben einen signifikant höheren ObjectLoss als die unveränderten Objekte ($\alpha = 0.05$). Zudem nimmt die Effektstärke d mindestens einen Wert von 0.80 ein (was lt. Bortz & Döring, 2002 der unteren Grenze für einen großen Effekt entspricht).

Kriterium 2: Die in der Proximitätenmatrix veränderten (= „nicht-gewussten“) Objekte haben einen signifikant höheren ObjectLoss als die unveränderten Objekte ($\alpha = 0.05$).

Ist Kriterium 1 erfüllt, dann ist das Verfahren technisch in der Lage, die geforderte Diagnoseleistung in ausreichendem Masse zu erbringen. Ist selbst Kriterium 2 nicht mehr erfüllt, dann gilt die Grundstruktur in der veränderten Matrix als zusammengebrochen. Die NMDS-Karte hat dann ihre Stabilität vollständig verloren.

4.2.2. Resultate

Abbildung 30 zeigt die ObjectLosses aller bekannten und aller unbekannten Objekte. Bei der unverfälschten Ausgangsmatrix sind die Abweichungen jeweils 0, weil ja identische Karten aufeinander gelegt werden. Bereits bei einem einzigen „nicht-gewussten“ Land verändert sich die NMDS-Karte, und zwar nicht nur für das veränderte Objekt, sondern auch für alle anderen. Dieser Effekt (ObjectLosses der gewussten Länder > 0) fällt umso größer aus, je weniger Objekte skaliert wurden: Bei 12 und 14 Objekten beträgt der ObjectLoss ca. 0.30, ab 20 Objekten fällt er dann aber auf vernachlässigbar kleine Werte unter 0.10 (Damit einher geht auch eine markante Verringerung der Standardabweichungen der ObjectLosses der gewussten Länder.). Dieses Muster in den Resultaten lässt sich darauf zurückführen, dass mit zunehmender Größe der Matrix die Robustheit des gewählten NMDS-Algorithmus greift und der Prokrustes-Transformation somit eine saubere Unterscheidung ermöglicht.

Wichtiger aber ist der Haupteffekt: Die nicht-gewussten Länder erhalten in der Prokrustes-Transformation jeweils den weit größeren ObjectLoss. Dieser liegt – relativ unabhängig von der Größe der Matrix – bei ca. 0.90. Diese Werte bleiben erstaunlich konstant, gleich wie viele Länder als nicht-gewusst gesetzt werden. Lediglich bei den großen Matrizen ergibt sich ein kleiner Aufwärtstrend von 0.80 bis 1.00, wobei aber der Mittelwert ebenfalls in der Region von 0.90 zu liegen kommt.

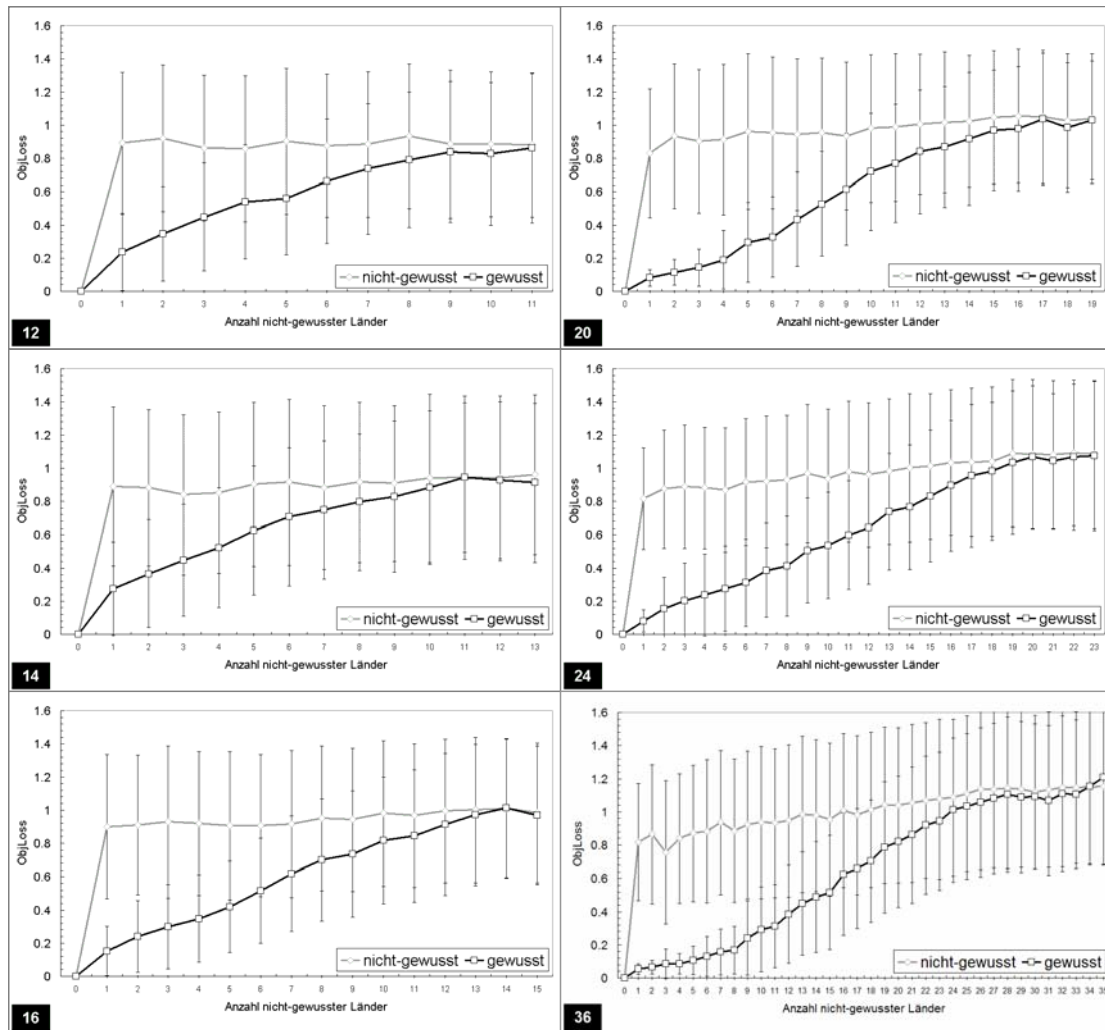


Abbildung 30. Gemittelte ObjectLosses aus jeweils 100 Simulationsdurchgängen. Die Zahlen in der linken unteren Ecke der einzelnen Verlaufsdiagramme bezeichnen die Objektmengen $n = 12, 14, 16, 20, 24, 36$. (Jeweils die letzte Kategorie auf der x-Achse steht für die vollständig aus Zufallszahlen bestehende Ähnlichkeitsmatrix. Diese Kategorie ist nicht gleich der Anzahl Objekte, da bei einer Datenform von paarweisen Ähnlichkeitsurteilen bereits bei $n-1$ alle Zellen in der Matrix verändert sind.)

Weitaus stärker ist dieser Aufwärtstrend bei den gewussten Objekten. Mit zunehmender Zahl ungewusster Länder steigen die ObjectLosses der gewussten markant an. Unabhängig von der Matrixgröße ist nach drei Viertel nicht-gewusster Objekte ein Niveau erreicht, das eine Unterscheidung zwischen gewussten und ungewussten Ländern auf der Ebene der Mittelwerte schwierig macht.

Zur Einordnung der Größe der gefundenen Mittelwertsunterschiede hatten wir zwei Kriterien formuliert: Das strengere Kriterium 1 legt die Grenze fest, bis zu der das Verfahren technisch brauchbare Vorhersagen macht, ob ein Land als gewusst oder als nicht-gewusst einzustufen ist. Für dieses Kriterium muss neben der Signifikanz im t -Test auch eine Effektstärke von $d > .80$ gegeben sein. Abbildung 31 zeigt, dass das bei 12 und bei 14 skalierten Objekten der Fall ist, solange nicht mehr als vier bis fünf Länder als nicht-gewusst gesetzt wurden. Für 16 Objekte dürfen 6 Länder nicht-gewusst sein, für 20 Objekte 8 Länder, für 24 Objekte 11 Länder und für 36 Objekte 16 Länder. Diese Zahlenverhältnisse pendeln also um einen Gesamtwert

von ca. 40% nicht-gewusster Länder, die für das vorgeschlagene Verfahren das technische Maximum darstellt, um in der Praxis brauchbar zu sein.

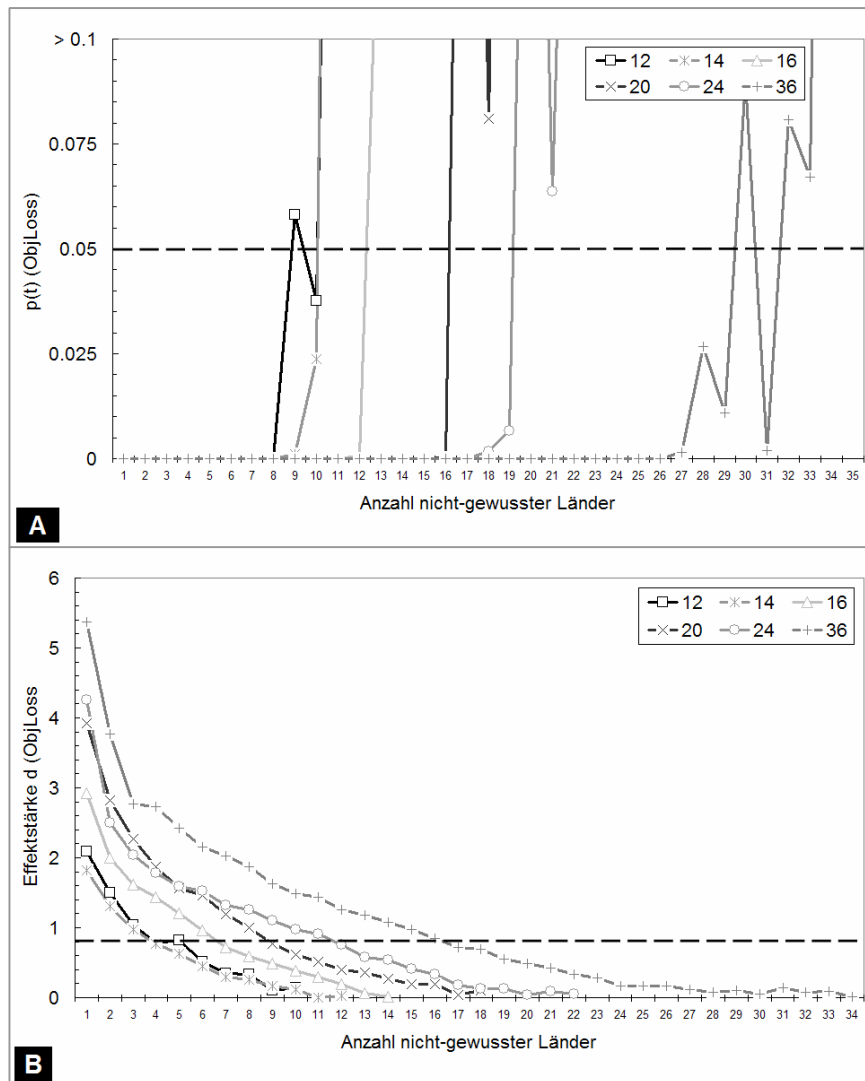


Abbildung 31. Signifikanz und Effektstärke der Mittelwertsunterschiede aus Abbildung 30. A) t-Tests (unabhängig, einseitig). Als gestrichelte Linie ist das 5%-Signifikanzniveau eingetragen B) Effektstärke d . Die gestrichelte Linie bei einem d von 0.8 steht für das Niveau einer hohen Effektstärke (Bortz, 2002, S. 604).

Interessant ist auch derjenige Punkt, an dem die Grundstruktur in einer Karte völlig zusammenbricht. Dies ist nach Kriterium 2 dann der Fall, wenn sich die ObjectLosses veränderter und nicht-veränderter Objekte nicht mehr signifikant voneinander unterscheiden. Das ist erst überraschend spät der Fall: Bei 12 Objekten müssen 9 verändert sein, bei 36 Objekten deren 28. Im Mittel pendeln die Werte, die noch zu einem signifikanten Unterschied führen, um die 75%-Marke. Dieser Befund über die Robustheit ist erstaunlich, denn bei drei Viertel veränderter Objekte sind ja bereits 15 von 16 Matrixwerten durch Zufallszahlen ersetzt! Damit spricht dieses Resultat für die enorme Robustheit der eingesetzten NMDS- und Prokrustes-Algorithmen.

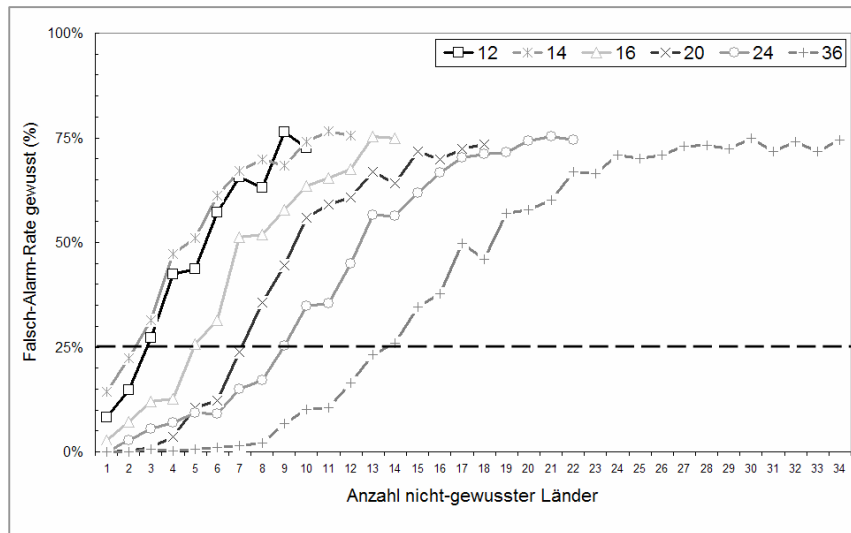


Abbildung 32. Falsch-Alarm-Rate der bekannten Objekte beim Kriterium einer Sensitivität von 75% für die nicht-gewussten Objekte. Die gestrichelte Linie steht für das Kriterium von 25% falscher Vorhersagen bei bekannten Objekten.

Im Hinblick auf die praktische Brauchbarkeit des getesteten Verfahrens ist eine detailliertere Analyse der Effektstärke angebracht. Bei einem automatisierten Einsatz in einem Intelligenten Tutoriellen System beispielsweise besteht die Aufgabe darin, möglichst viele nicht-gewusste Objekte zu detektieren, ohne dabei allzu häufig „falschen Alarm“ auszulösen. Da Intelligente Tutorielle Systeme in der Regel in zyklischen Schleifen ihre Lektionen aufbauen, bedarf es bei der Diagnostik keiner Sensitivität von 100%: Was beim ersten Mal nicht erfasst wurde, kann auch später noch in die Lektionen einfließen. Aus diesem Grund wird für die in Abbildung 32 präsentierte Analyse eine Sensitivität von 75% zugrunde gelegt. Legen wir das praktische Brauchbarkeitskriterium für die Falsch-Alarm-Rate auf 25%, dann zeigt sich, dass bei einem Gesamtumfang von 12 und 14 Objekten bereits bei drei nicht-gewussten Ländern mehr als 25% der gewussten Länder (also 3 von 9 respektive 11) ebenfalls als nicht-gewusst diagnostiziert werden. Bei 16 Objekten ist die Grenze bei 6 Objekten überschritten, bei 20 Objekten bei 8 Ländern, bei 24 Objekten bei 10 und bei 36 Objekten bei 14 Ländern. Verallgemeinert lässt sich sagen, dass bei Karten einer Größe von 16 oder mehr Objekten ab 40% nicht-gewusster Länder die Falsch-Alarm-Rate zu hoch wird, als dass das Verfahren hinlänglich gute automatisierte Empfehlungen für die Auswahl von Objekten für die nächsten Lektionen zu geben in der Lage wäre.

4.2.3. Diskussion

Die Simulation hat gezeigt, dass das vorgeschlagene Verfahren ausgesprochen robust und damit technisch in der Lage ist, gewusste Objekte (mit konsistent zueinander passenden Ähnlichkeitsurteilen) von nicht-gewussten Objekten (mit zufällig verfremdeten Ähnlichkeitsurteilen) zu unterscheiden. Der Unterschied in den gemittelten ObjectLosses ist vor allem bei wenigen nicht-gewussten Objekten enorm und bleibt bis zur 40%-Marke in der Region eines großen Effektes. Erst dann wird der Effekt zu klein, als dass damit eine Praxisanwendung unbeschadet und in jedem Fall sinnvoll wäre.

Die Analyse der Falsch-Alarm-Raten bei gegebener Sensitivität von 75% ergibt ebenfalls eine Grenze von ca. 40% der Objektzahl, ab der eine automatisierte Auswahl von Objekten nicht mehr vernünftig greifen würde. Daraus folgt die Empfehlung, dass dieses Verfahren erst dann den Einsatz lohnen könnte, wenn bereits von einem individuellen Vorwissen von 60% oder mehr ausgegangen werden kann, wenn aber dem System nicht bekannt ist, auf welche Objekte sich dieses Vorwissen bezieht.

Dabei erweist sich die Funktionsfähigkeit prozentual betrachtet als relativ unabhängig von der Anzahl der im Raum zu skalierenden Objekte. Nachdem Läge et al. (2005) gezeigt haben, dass Robustheit eine direkte Funktion der Matrixgröße ist, war davon nicht unbedingt auszugehen. Allerdings muss darauf hingewiesen werden, dass die Simulation von einer praktisch ausreißerfreien Proximitätsmatrix als Normmodell ausging. Bei weniger als 16 zu skalierenden Objekten können einige wenige Ausreißer in den Ähnlichkeitsurteilen aber die Karte bereits beeinflussen. Dem korrespondiert das Resultat der Simulation, dass bei nur einem einzigen nicht-gewussten Land der mittlere ObjectLoss der unveränderten Objekte bereits auf einen Wert von 0.30 springt, während er ab 16 Objekten deutlich niedriger ausfällt. Deswegen bleibt – trotz der positiven Befunde aus der Simulation – bei weniger als 16 Objekten in der Wissenskarte Vorsicht geboten, und zwar gerade im Hinblick auf ein vollständig automatisiertes Verfahren.

Außerdem zeigen die Standardabweichungen in Abbildung 30, die Effektstärken in Abbildung 31 und die Falsch-Alarm-Raten in Abbildung 32, dass bei einer relativ kleinen Anzahl nicht-gewusster Objekte die Diagnose weitaus besser gelingt, wenn die Gesamtzahl der zu skalierenden Objekte größer ist. Interessanterweise besteht hier kein nennenswerter Unterschied zwischen 20 und 24 Objekten (die Streuung der ObjectLosses gewusster Länder ist bei 20 Objekten sogar kleiner!), wohl aber einer zwischen 16 und 20 Objekten. Als Optimierung von Genauigkeit und Ökonomie leitet sich zumindest aus den hier vorliegenden Resultaten die Empfehlung ab, für die Diagnose einiger weniger nicht-gewusster Objekte (laut Simulationsergebnissen bis zu 7) eine insgesamt zu skalierende Anzahl von 20 Objekten zu verwenden.

4.3. Experiment

4.3.1. Methode

Die Simulationsstudie hat gezeigt, dass das vorgeschlagene Verfahren grundsätzlich in der Lage ist, gewusste von nicht-gewussten Objekten zu unterscheiden. Allerdings wurden dabei hinsichtlich der Ähnlichkeitsurteile zwei idealisierende Annahmen gemacht: Im Fall von Wissen sind die Ähnlichkeitsurteile konsistent und gut kalibriert, und im Fall von Nicht-Wissen sind sie chaotisch. Diese Annahmen stellen sozusagen den „besten Fall“ für das vorgeschlagene Verfahren dar. Im folgenden Experiment soll nun überprüft werden, wie nahe die Vorhersagegüte bei Ähnlichkeitsurteilen lernender Personen an die Ergebnisse der Simulationsstudie herankommt.

Durchführung. Eine repräsentative Auswahl von 20 mitteleuropäischen Singvogelarten bildete das Wissensgebiet für das Lernexperiment. Diese Auswahl wurde in Zusammenarbeit mit Ornithologie-Experten getroffen. Jede Vogelart ist durch ihre Ausprägungen in den Merkmalen Aussehen, Familienzugehörigkeit, Habitat, Nahrung, Schnabelform, Nest/Nistplatz, Gesang (verbal und auditiv) und Brutstatus beschrieben. Die Informationen stützen sich dabei auf gängige Einführungswerke und Feldführer (Bezzel, 1985; Hudec, 1992; Johnson, 1992 sowie United Soft Media, 2001). Sechs Ornithologen gaben jeweils 190 paarweise Ähnlichkeitsurteile zwischen allen 20 Vogelarten ab. Die gemittelte Proximitätenmatrix wurde durch NMDS in eine zweidimensionale Karte umgesetzt, die als Normkarte für das Lernexperiment diente (Abbildung 33).

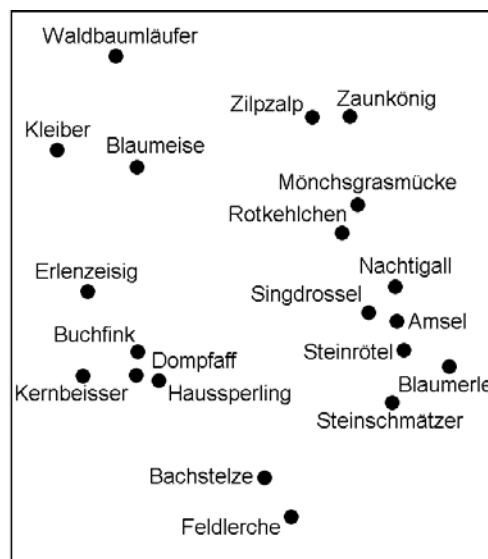


Abbildung 33. Expertenstruktur Singvögel (NMDS-Karte auf Basis der gemittelten Ähnlichkeitsurteile von sechs Ornithologen).

Die Expertenkarte zeigt eine ringförmige Verteilung der ausgewählten Arten über den Raum (d.h. das Zentrum der Karte bleibt unbesetzt). Auf der linken Seite befindet sich das Finken-Cluster, in dem die typischen Körnerfresser (Kernbeißer, Dompfaff, Buchfink und, wegen seines feineren Körperbaus etwas zu den Meisen hin verschoben, der Erlenzeisig) liegen. Diesem Cluster ist der Haussperling (ebenfalls vornehmlich von Körnern und Sämereien lebend) zugeordnet. Die Arten waren von vornherein so ausgewählt worden, dass allein aus den Namen keinerlei Rückschlüsse auf die Clusterzugehörigkeit möglich sind. Das gilt auch für das Drosselcluster auf der rechten Seite (Steinschmätzer, dann die engen Verwandten Blaumerle und Steinrötel, die ebenfalls eng verwandten Amsel und Singdrossel, sowie die als Paar ebenfalls verwandtschaftlich zusammengehörenden Rotkehlchen und Nachtigall). Den Drosseln ist gemeinsam, dass sie vornehmlich von Bodeninsekten und anderen Kleintieren am Boden sowie von Früchten und Beeren leben. Dieser Unterschied in der Ernährung stellt eine wesentliche Bedingung für Körperbau und Lebensweise dar, weshalb die Experten durchgängig Finken und Drosseln einander in der Karte gegenüberstellen.

Im oberen Teil der Karte befinden sich Singvögel, die vornehmlich von Insekten leben und sich in eher geschlossenen Lebensräumen aufhalten (Hecken oder Wald). Dabei gehören

Mönchsgrasmücke und Zilpzalp als sog. „Laubsänger“ zusammen, wobei erstere den kleinen Drosseln (Rotkehlchen und Nachtigall) in Art und Habitat sehr ähnlich ist und deswegen etwas weiter in diese Richtung platziert wird. Der Zaunkönig bildet den Abschluss dieser Busch- und Heckenbewohnenden Gruppe (am „südlichen“ Ende des Drosselclusters stehen mit Steinrötel, Blaumerle und Steinschmätzer eher die Vertreter offenerer Habitate).

Unten in der Karte befinden sich mit Feldlerche und Bachstelze zwei ebenfalls verwandte und hauptsächlich auf Insektennahrung spezialisierte Arten, die offenes Gelände bevorzugen und deswegen gut hierher passen. Deren Antipode ist der Waldbaumläufer, welcher sein geschlossenes Habitat nicht verlässt. Unter ihm gruppieren sich Kleiber (der wie der Baumläufer vertikal entlang der Baumstämme läuft) und die Blaumeise. Beide gehören den Meisenartigen an (der Kleiber wird auch „Spechtmeise“ genannt). Obwohl sie eher feine, pinzettenförmige Schnäbel haben, werden sie von den Experten in die Nähe der Finken gerückt, weil die Meisen als Standvögel im Winter auf Körner und Sämereien ausweichen müssen und dann gemeinsam mit den Finken den Lebensraum teilen.

An diese Expertenkarte sollten sich die Versuchspersonen in einem mehrstufigen Experiment Objekt für Objekt annähern. Sie hatten die dazu Aufgabe, sich anhand von den oben genannten Merkmalsdimensionen ein Basiswissen zu den 20 Singvogelarten anzueignen. Die Lerner nahmen, verteilt auf fünf aufeinander folgende Wochen, an fünf Sitzungen zu je eineinhalb Stunden teil (Abbildung 34). Nachdem aus der Simulationsstudie zu folgern war, dass ein Vorwissen über eine gewisse Anzahl von Objekten technisch notwendig ist, damit überhaupt eine stabile Wissenskarte entsteht, dienten die ersten beiden Sitzungen der Bildung einer Grundstruktur von 10 Vogelarten. Insgesamt wurden diese ersten zehn Objekte drei Mal gelernt.

Dieses Wissen ermöglichte es den Lernern, den Wissensraum aufzubauen und eine kognitive (Basis-)Struktur zu entwickeln. Dazu wurden 10 Vogelarten so ausgewählt, dass eine neue NMDS-Skalierung dieser 10 Objekte strukturell möglichst nahe an den Wissensraum von 20 Objekten kam (in Abbildung 35 wird bei den Resultaten dazu ein Beispiel gegeben). Die auf diese Weise entstehende „Grundstruktur“ hatte die 10 Objekte dann bereits in der korrekten Ordnung zueinander, d.h. die Versuchspersonen brauchten hinsichtlich der zunächst gelernten 10 Objekte im weiteren Verlauf des Experimentes nicht mehr zu akkomodieren (vgl. dazu die Lernempfehlungen von Ausubel, 1963, 2000 sowie die Befunde zu Akkomodation und Assimilation in Wissenskarten in Kapitel 3 in dieser Arbeit).

Ab der dritten bis zur fünften Sitzung lernten die Versuchspersonen schrittweise zehn weitere Objekte in Zweierpaketen (insgesamt 5 Lernphasen). Auf diese Weise wurden insgesamt 6 verschiedene Wissensstände pro Versuchsperson kreiert: 10 von 20 Objekten gewusst, 12 von 20 gewusst, 14, 16, 18 sowie 20 von 20 Objekten gewusst. Die zu jedem Wissensstand durchgeführte Diagnose durch Ähnlichkeitsurteile ermöglichte dann für die Auswertung eine Überprüfung, inwieweit die gewussten / nicht-gewussten Objekte vom vorgeschlagenen System detektiert wurden.

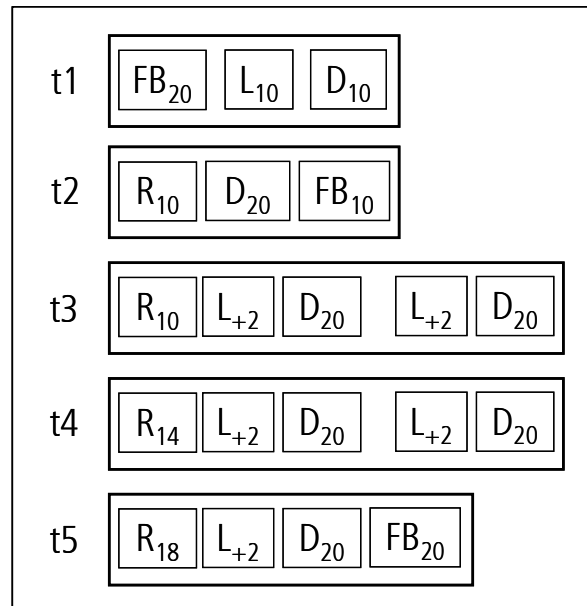


Abbildung 34. Experimentalsetting. t1-t5 bezeichnen die fünf Sitzungen, verteilt auf fünf aufeinander folgende Wochen; FB = Fragebogen; L = Lernen; D = Ähnlichkeitsdiagnose; R = Repetition des bereits Gelernten. Indizes bezeichnen die Anzahl Objekte, welche getestet bzw. gelernt wurden.

Das Lernen in jeder zehnminütigen Lernphase fand selbstgesteuert am Computer statt, wobei über eine Menüliste für jeden Vogel alle jeweiligen Merkmalsinformationen aufgerufen werden konnten. Zu Beginn einer Sitzung wurden jeweils die bereits gelernten Objekte repetiert. Zwischen den Lernphasen innerhalb einer Sitzung gab es eine Pause von fünf Minuten.

Anschließend an jede Lernphase erfolgte eine Statusdiagnose: Ähnlichkeitsurteile zwischen allen 190 Objektpaaren dienten als Datengrundlage zur Messung der Wissensstruktur der Lerner. Die Erhebung aller Paarvergleiche fand auf einer Skala von 1 bis 9 (1 = niedrigste Ähnlichkeit; 9 = höchste Ähnlichkeit) statt, wobei die 190 Urteile auf der Basis der gelernten Merkmalsdimensionen abzugeben waren.

Das Experimentalsetting macht die optimistische Annahme, dass die gelernten Merkmalsinformationen von der Versuchsperson auch gewusst werden. Denn nur dann kann sie sie ja in die Ähnlichkeitsurteile einfließen lassen. Gilt diese Annahme nicht, so arbeitet das gegen die Hypothese der Funktionstauglichkeit des von uns vorgeschlagenen Verfahrens, denn die Auswertung würde dann für jedes gelernte, aber de facto doch nicht gewusste Objekt einen „falschen Alarm“ anzeigen.

Um das zu vermeiden, wäre es notwendig, nach jeder Ähnlichkeitsdiagnose einen Wissensfragebogen einzusetzen, um überprüfen zu können, wie gut die Versuchsperson denn nun tatsächlich über jedes Objekt Bescheid weiß. Das wäre aber derart unökonomisch und für die (freiwilligen) Versuchspersonen unzumutbar, dass darauf verzichtet werden musste. Als eine Art Kompromiss wurde zu Beginn des Experiments mit einem hierarchisch gestuften Multiple-Choice-Fragebogen das bereits vorhandene Vorwissen getestet. Am Schluss der zweiten und am Schluss der fünften Sitzung wurde zusätzlich zu den Ähnlichkeitsurteilen ebenfalls das Merkmalswissen über diesen Fragebogen erhoben. Seine hierarchische Stufung sah auf der ersten Ebene eine Einteilung der Objekte in grobe Merkmalskategorien vor (z.B. Merkmal Habitat: Kategorie 1 = „Wald sowie Unterholz, Hecken und Sträucher“; Kategorie 2 = „offe-

nes Gelände“; usw.). Auf der zweiten Ebene galt es dann, die Objekte detailliert beschriebenen Kategorien zuzuordnen (z.B. Merkmal Habitat: Kategorie 1.1 = „Kommt überall vor, wo es Bäume gibt. Laub- und Nadelwald, Parks, Gärten und Obstgärten“). Letztere waren dieselben Formulierungen der Merkmalsausprägungen wie die gelernte Textform.

Stichprobe. 16 Versuchspersonen (8 Frauen, 8 Männer) im Alter von 21 bis 45 Jahre nahmen an diesem Experiment teil. Da alle Teilnehmenden Studierende der Universität Zürich sind oder über eine höhere Berufsausbildung verfügen, war die formale Grundbildung relativ homogen. Alle Versuchspersonen waren zudem Laien in Sachen Ornithologie.

Datenauswertung und erwartete Befunde. Sämtliche Auswertungen der Daten fanden auf individuellem Niveau statt, und zwar analog zur Auswertung der Simulationsstudie: Die Dreiecksmatrizen der paarweisen Ähnlichkeitsurteile wurden mit dem NMDS-Algorithmus RobuScal (Läge et al., 2005) zweidimensional skaliert. Durch Prokrustes-Transformation wurden dann die individuellen Karten auf das Expertenmodell gelegt. So konnte in jeder Karte für jede Vogelart der ObjectLoss bestimmt werden.

Der erwartete Befund ist, dass die jeweils noch nicht gelernten Arten einen größeren ObjectLoss aufweisen als die bereits gelernten. Dabei ergeben sich bei den Daten jeder Versuchsperson sequentiell die Zahlenverhältnisse 10 Gelernte zu 10 Nichtgelernten (= 50% gelernt), 12 gelernte zu 8 nicht gelernten (= 60% gelernt), 14 gelernte zu 6 nicht gelernten (= 70% gelernt), 16 gelernte zu 4 nicht gelernten (= 80% gelernt) und 18 gelernte zu 2 nicht gelernten (= 90% gelernt). Je größer der Anteil gelernter Objekte ist, umso stabiler sollte die Wissenskarte einer Versuchsperson sein, und umso klarer sollte dementsprechend auch der Unterschied im ObjectLoss hervortreten. Wenn es ein Zahlenverhältnis zwischen gelernten und nichtgelernten Objekten gibt, ab dem sich die ObjectLosses nicht mehr systematisch voneinander unterscheiden (auch hier setzen wir neben Signifikanz im *t*-Test eine Effektstärke von $d = 0.80$ als strenges Kriterium an), dann ist die kritische Grenze für den Anteil notwendigen Vorwissens zu ziehen, damit die vorgeschlagene Diagnostikmethode greift.

Die maximal zu erwartende Effektgröße ist durch die Simulationsergebnisse festgelegt – sie basieren ja auf der Annahme, dass zwischen je zwei gewussten Objekten korrekte und in sich konsistente Ähnlichkeitsurteile abgegeben werden. Für das experimentelle Ergebnis sind daher Abstriche zu erwarten. Außerdem darf nicht davon ausgegangen werden, dass sämtliche Informationen aus den Lektionen auch von einer Versuchsperson gelernt wurden. Es kann also sein, dass die Auswertung einen Vogel als „gewusst“ registriert, von dem die Versuchsperson in Wirklichkeit noch nicht genug weiß. Das würde zu einer erhöhten Falsch-Alarm-Rate führen.

Voraussetzung für eine sinnvolle Überprüfung der Hauptfragestellung ist also, dass die Versuchspersonen in dem Experiment überhaupt hinreichend viel über die Objekte gelernt haben. Das ist an drei Befunden zu überprüfen: (1) Die Probanden haben nach dem Lernen der ersten 10 Objekte eine mit der Expertenkarte vergleichbare Grundstruktur in ihren Köpfen. (2) Die Verschiebungen der Punktekongfigurationen der Lerner beim sequentiellen Dazulernen der neuen Objekte laufen am Ende des Experiments zielgerichtet auf die Expertenkarte zu. (3)

Die sequentiellen Veränderungen der Lerner-Karten gehen mit einer Zunahme an Merkmalswissen einher und sind daher auf diese Wissenszunahme zurückzuführen.

Und schließlich lässt sich erwarten, dass die Leistung des Diagnoseverfahrens ganz maßgeblich davon abhängt, wie gut die Grundstruktur bei der einzelnen Person etabliert ist: Bei einem Lerner, der einen Grossteil der gelernten Objekte kennt und gut zu integrieren weiß, sollte der Unterschied im ObjectLoss zwischen gewussten und nicht-gewussten Vogelarten größer ausfallen als bei einer Person, die die gelernten Objekte noch nicht so gut in Relation zueinander zu stellen vermag.

Um dies zu überprüfen, wird die Stichprobe der 16 Versuchspersonen in zwei Hälften eingeteilt, und zwar jeweils anhand der Übereinstimmung der individuellen Grundstruktur mit dem Expertenmodell für die Vogelarten der Grundstruktur. Dazu werden die abgegebenen Ähnlichkeitsurteile für die ersten 10 gelernten Objekte aus der jeweiligen Matrix einer Person extrahiert und einer NMDS unterzogen. Diese Karte wird per Prokrustes-Transformation auf diejenige NMDS-Karte gelegt, die sich ergibt, wenn man die entsprechenden Ähnlichkeitsurteile aus der Expertenmatrix extrahiert. Der AverageLoss dieser beiden Karten gibt dann an, wie gut das Grundwissen bei einer jeden Person gelernt und integriert ist – und das völlig unabhängig davon, wie sie alle anderen Objekte kennt und behandelt.

Die resultierende Hypothese lautet, dass für die Untergruppe der Personen mit besser integriertem Vorwissen der Unterschied der ObjectLosses (und damit auch die Effektstärke) zwischen gelernten und nicht-gelernten Objekten größer ausfällt als bei der Untergruppe der Personen mit weniger gut integriertem Vorwissen.

4.3.2. Resultate

Die Basiskarte von 10 Objekten. Unmittelbar nach dem zweiten Erlernen der 10 Objekte der Basisstruktur gaben die Versuchspersonen in der zweiten Sitzung Ähnlichkeitsurteile über alle 20 Objekte ab. Aus dieser Diagnose ließ sich die Teilmatrix für die 10 bereits gelernten Objekte extrahieren und mit NMDS skalieren. Die resultierenden individuellen Wissenskarten wurden dann mittels Prokrustes-Transformation auf die Expertenstruktur gelegt. Die gemittelten AverageLoss-Werte der Versuchspersonen betragen 0.46. Das darf als hinreichende Annäherung an die Expertenstruktur gesehen werden, um davon zu sprechen, dass die Probanden durch das Merkmalslernen eine Grundstruktur der zehn Objekte aufgebaut haben. Die Versuchspersonen halten dieses AverageLoss-Niveau für die Teilkarte der zehn Objekte über das ganze weitere Experiment hin im Mittel konstant. Abbildung 35 zeigt ein Beispiel für eine durchschnittlich gute individuelle Lernerkarte.

(Da für die Abschätzung der Qualität einer Prokrustes-Transformation zufallskritische AverageLoss-Grenzwerte ohne praktische Aussagekraft sind, soll die gerade getroffene Einschätzung durch zwei Referenzwerte zur Einordnung des Ergebnisses untermauert werden: (1) Die individuellen Expertenkarten weichen mit einem durchschnittlichen AverageLoss von 0.34 von der gemittelten Expertengrundstruktur ab. (2) In einem Lern-Experiment (vgl. Kapitel 2 in dieser Arbeit) erzielten die Probanden am Ende eines Lernexperimentes, ebenfalls zur Ornithologie, Abweichungswerte um die 0.40.)

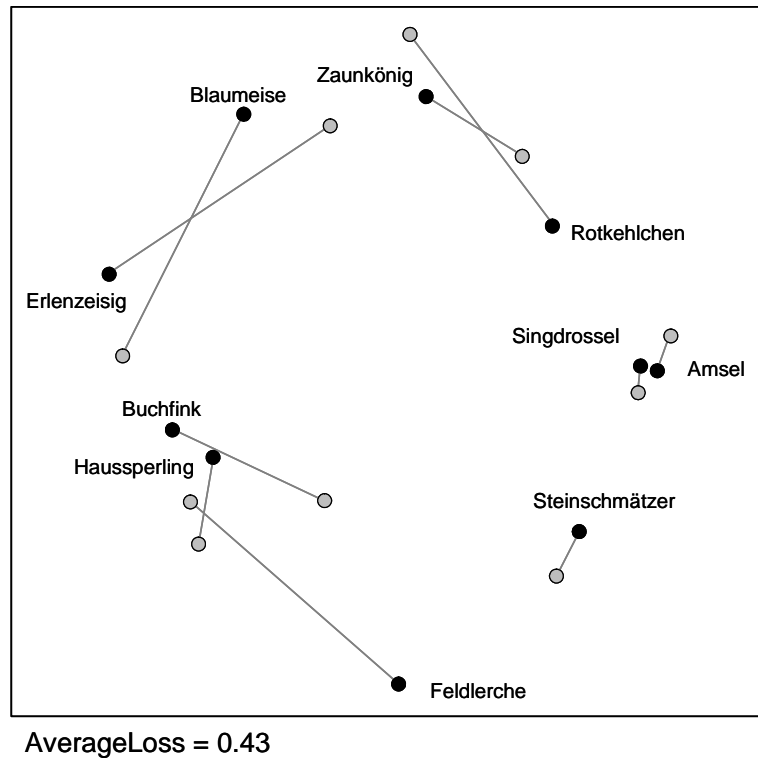


Abbildung 35. Prokrustes-Transformation der NMDS-Lösung der Grundstruktur von Versuchsperson 8 (die eine durchschnittliche Leistung erzielte: AverageLoss = 0.43) mit dem Expertenmodell (schwarz). Eine grundsätzliche Strukturähnlichkeit der Karten wird deutlich. Allerdings erkannte die Versuchsperson nicht die Nähe des Rotkehlchens zu den Drosseln, und die Bildung des Finkenclusters ist noch nicht vollzogen.

Konvergenz der Wissenskarten am Ende des Experiments. Wenn die Probanden alle 20 Objekte gelernt haben, sollten die individuellen Wissenskarten nach der letzten Lernphase hinlänglich nah an die Expertenkarte herankommen. Natürlich kann in einer experimentellen Studie mit Laien in der begrenzten Lernzeit nicht erwartet werden, dass die Karten genauso gut werden wie die von langjährigen Experten. Aber für die Auswertung der Hauptfragestellung muss die Annäherung groß genug ausfallen, dass von einer systematischen Verbesserung gegenüber dem Wissensstand vor dem Dazulernen der zweiten zehn Vogelarten gesprochen werden kann. Der gemittelte AverageLoss-Wert der Versuchspersonen nach dem Vorlernen (10 Objekte bekannt, 10 Objekte unbekannt) beträgt für die Karte mit allen 20 Objekten 0.69. Dieser Werte nimmt über den Lernverlauf markant ab und beträgt zum Schluss (alle Objekte bekannt) noch 0.49 ($t(15) = 4.51, p < 0.001$). Als Referenzwert für die Einschätzung der Größe der Verbesserung mag die Information dienen, dass die individuellen Expertenkarten zur gemittelten Expertenkarte einen durchschnittlichen AverageLoss von 0.40 aufweisen.

Zunahme des Merkmalswissens. Die unmittelbare Messung des Merkmalswissens wurde in den Sitzungen 1 (Vorwissen), 2 (gelerntes Wissen über 10 Objekte) und 5 (gelerntes Wissen über 20 Objekte) jeweils über hierarchisch gestufte Fragebögen hohen Schwierigkeitsgrades durchgeführt. Die durchschnittlich erreichte Punktzahl steigt von t1 nach t5 von 60 auf 101 Punkte (maximal mögliche Punktzahl wäre 200). Dieser Wissenszuwachs ist im t -Test signifikant ($t(15) = -15.23, p < 0.001$). Auch für die 10 gelernten Objekte (max. mögliche Punktzahl = 100) ist der Wissenszuwachs signifikant und steigt von t1 nach t2 von 33 auf 64 Punkte

($t(15) = -10.87, p < 0.001$). Damit kann die Grundvoraussetzung, nämlich dass durch die Lernphasen das Wissen verbessert wurde, als gegeben betrachtet werden.

Die Absolutwerte zeigen jedoch an, dass der Lernerfolg weit von einem perfekten Grad an Expertise entfernt ist. Demnach ist zu schlussfolgern, dass längst nicht jede gelernte Vogelart auch gleich zu einer „gewussten“ Vogelart wird. Da dieser Effekt seiner Richtung nach asymmetrisch ist (gelernte Objekte können nicht-gewusst bleiben, nicht-gelernte hingegen nicht zu gewussten werden), wird die Messung der Diskriminationsleistung der ObjectLosses in diesem Experiment tendenziell zu einer systematischen Unterschätzung der Diagnosefähigkeit des vorgeschlagenen Verfahrens führen.

Zusammenhang zwischen Merkmalswissen (Fragebogen) und ObjectLoss. Mit dem Konstatieren eines deutlichen Lernerfolgs in Fragebögen und Karten ist die notwendige Grundvoraussetzung gegeben, um bei den gemessenen ObjectLosses die prognostizierten Unterschiede zwischen gewussten und nicht-gewussten Vogelarten finden zu können. Dazu werfen wir zunächst einen Blick auf den korrelativen Zusammenhang zwischen Merkmalswissen und ObjectLoss. Wenn für eine Vogelart im Fragebogen ein hoher Punktwert erzielt wird, dann sollte sich das in den Karten in einem tiefen ObjectLoss äußern und umgekehrt. Zum Zeitpunkt t_2 ist genau die Hälfte der Objekte gelernt, so dass wir aufgrund des momentanen Wissensgefälles einen solchen Zusammenhang erwarten. In den Messungen der 5. Sitzung hingegen sind alle Objekte gelernt worden. Hier erwarten wir, dass Schwankungen in Fragebogen und ObjectLosses eher zufällig ausfallen, dass es also keine systematische Korrelation zwischen diesen beiden Messungen gibt. Die Auswertung bestätigt diese Vermutungen: Die Produkt-Moment-Korrelation zwischen Punktzahl pro Vogelart im Fragebogen und ObjectLoss aus der Prokrustes-Transformation beträgt nach dem Lernen von 10 Vogelarten $r = -0.49$ ($p < 0.05$, einseitig), nach der letzten Lernphase jedoch $r = 0.07$ ($p = 0.39$, einseitig).

Vorhersagegüte der ObjectLosses. Die ObjectLosses der noch nicht gelernten Vogelarten in den individuellen Lernerkarten sollen systematisch höher ausfallen als die ObjectLosses für bereits gelernte Arten. Abbildung 36A zeigt, dass das durchgehend der Fall ist. Für 10 – 16 gelernte Arten gehen die p -Werte der t -Tests gegen 0, nur bei 18 bekannten / 2 unbekannten Objekten ist der Unterschied lediglich auf dem 5%-Niveau signifikant (Abbildung 36B). Trotz des signifikanten Unterschieds spricht allerdings die Effektstärke d für 10 – 14 gelernte Objekte nur für einen mittelgroßen Effekt (vgl. Bortz & Döring, 2002, Cohen, 1988). Eine hohe Effektstärke wird nur dann erzielt, wenn 16 Objekte gelernt und 4 noch nicht gelernt sind.

Die Einordnung dieser Effektstärken hinsichtlich ihrer praktischen Bedeutung für die Zusammenstellung von Objekten für eine Lektion wird durch Abbildung 36C möglich. Genau wie in der Simulation, geht diese Auswertung von einer hinreichenden Sensitivität von 0.75 aus. Die Falsch-Alarm-Raten liegen über 50% und sind damit doch beträchtlich hoch. Für 10 nicht-gewusste Objekte ist dieser Wert gleich hoch wie in der Simulation. In dieser jedoch sinken die Werte bei 8 nicht-gewussten Objekten auf 35% und bei 6 nicht-gewussten Objekten auf 10%. Zwar sinkt auch im Experiment die Falsch-Alarm-Rate mit zunehmender Anzahl gelernter Objekte (wie wegen der größer werdenden Stabilität der Wissenskarte erwartet wur-

de), fällt jedoch selbst bei 16 gelernten Vogelarten nur auf einen Wert von 42% und steigt danach überraschenderweise sogar wieder an. Auch in den Kurven der Abbildung 36A und Abbildung 36B geht die Schere nicht wie erwartet mit zunehmender Zahl gelernter Objekte weiter auf, sondern bei 18 gelernten Objekten sogar wieder zu.

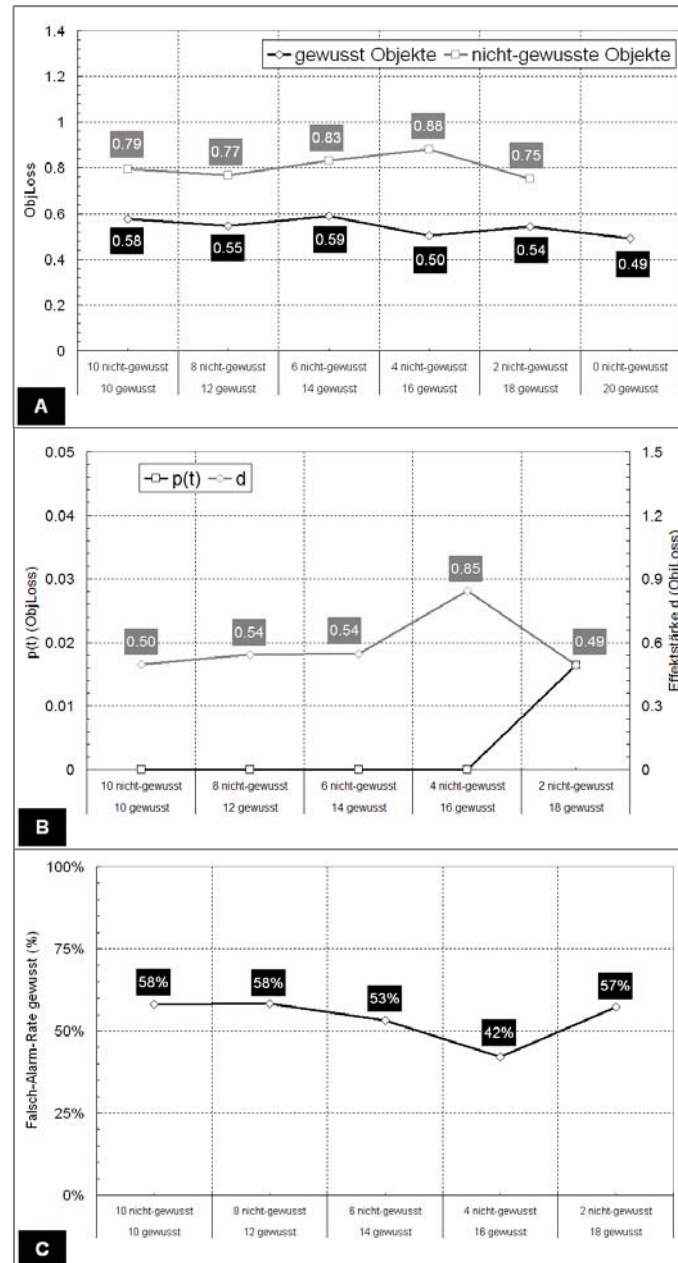


Abbildung 36. Objektunterschiede zwischen den Lernerkarten verglichen mit der Expertenkarte im Experiment. A) Mittelwerte der ObjLoss-Werte der gelernten und der nicht gelernten Objekte in Abhängigkeit des Wissensstandes. Der ObjLoss-Wert bei „0 nicht-gewusst“- und „20 gewusst“-Objekten steht für die mittlere Objektabweichung der Versuchspersonen zur Expertenkarte, nachdem alle Information mindestens einmal gelernt worden war (Referenzwert). B) Inferenzstatistische Analyse der Mittelwertsunterschiede mittels t-Test (unabhängig, einseitig) und Effektstärke d. C) Falsch-Alarm-Rate der bekannten Objekte beim Kriterium 25% Miss der unbekannten Objekte.

Eine alternative Form, die praktische Brauchbarkeit des ObjectLosses zur Detektion nicht-gewusster Objekte heranzuziehen, ergibt sich durch die Angabe, wie viele Objekte aufgrund

des ObjectLosses ausgewählt werden müssen, um pro Wissensstand (= Lernsitzung) zwei nicht-gewusste Objekte in der Auswahl zu haben. Diese Berechnung findet sich in Abbildung 37. Die Werte liegen alle weit unter dem Erwartungswert, den man ohne diese Diagnose hätte, aber sie liegen auch deutlich über dem Minimum von 2 (was jeweils eine perfekte Vorhersagbarkeit wäre). Wiederum fällt auf, dass der Wert für 18 gelernte Objekte inklusive der Nachtigall deutlich schlechter wird.

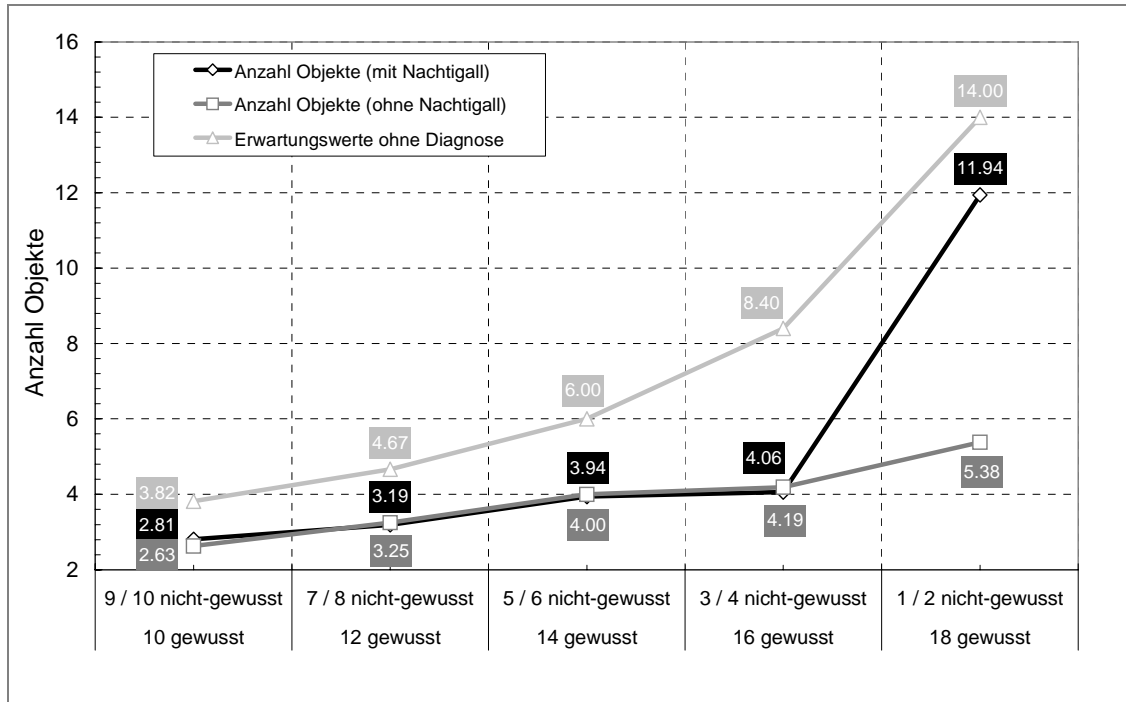


Abbildung 37. Gesamtzahl auszuwählender Objekte, um zwei nicht-gelernte dabei zu haben. Auswertung für alle 20 Objekte sowie Reanalyse ohne Nachtigall.

4.3.3. Diskussion

Wie erwartet bewegen sich die Befunde für die ObjectLosses im Korridor zwischen statistischer Signifikanz (als minimalem Anspruch) und der Vorhersagegüte aus der Simulation (als maximaler Leistungsgrenze). Für eine Einordnung dieses Ergebnisses ist zu berücksichtigen, dass die Versuchspersonen nicht jedes Objekt auch gewusst haben, sobald es in der Lerneinheit vorgekommen war (Insgesamt gesehen schnitten die Versuchspersonen im Merkmalsfragebogen deutlich schlechter ab, als wir das erwartet hätten. Allerdings war der Fragebogen durch sein gestuftes Verfahren auch recht anspruchsvoll.). Es muss also konzediert werden, dass so manches gelernte Objekt mit einem hohen ObjectLoss vermutlich in der Tat korrekt als „nicht-gewusst“ detektiert wurde.

Derartige Abstriche sind in einem aufwändigen Lernexperiment durchaus an der Tagesordnung. Was allerdings ins Auge springt, ist die Verschlechterung der Diagnoseleistung bei nur noch zwei ausstehenden Vogelarten. Zu erwarten war ja, analog zur Simulationsstudie, auf jeden Fall eine stetige Zunahme der Funktionstüchtigkeit des vorgeschlagenen Verfahrens: Die Prokrustes-Transformation sollte falsch platzierte Objekte umso deutlicher als solche markieren können, je größer der Anteil gut platzierter Objekte ist. Vor allem in Abbildung

36B ist aber deutlich zu sehen, dass die Effektstärke d bei zwei unbekannten Objekten ($d = 0.49$) sogar etwas niedriger ist als bei zehn unbekannten ($d = 0.50$). Auch in den Abbildung 36A und Abbildung 36C läuft der Wert für 18 bekannte / 2 unbekannte Objekte damit gegen den eigentlich zu erwartenden Trend.

Eine genauere Reanalyse auf der Basis aller Einzelobjekte brachte dann einen überraschenden Befund: Die Nachtigall, eine der beiden von allen Versuchspersonen zuletzt gelernten Vogelarten, wurde konsequent von der Mehrzahl der Probanden bei allen Ähnlichkeitsdiagnosen in die adäquate Region der Karte gebracht und erhielt damit systematisch kleine ObjectLosses (Der über alle Diagnosezeitpunkte gemittelte ObjectLoss der Nachtigall beträgt 0.57 und liegt somit genau auf dem Niveau der bekannten Objekte!). Diese gute Qualität der abgegebenen Ähnlichkeitsurteile zur Nachtigall verblüfft, denn sie findet in den Punktzahlen der Fragebögen überhaupt keine Entsprechung: Die Personen verfügen über kein besseres Merkmalswissen zur Nachtigall als zu den anderen noch nicht gelernten Vogelarten. Und trotzdem geben sie ihre Ähnlichkeitsurteile so ab, als hätten sie die Merkmale dieses Vogels bereits gelernt.

Ein derartiger systematischer Effekt senkt natürlich die Leistung des vorgeschlagenen Diagnostikverfahrens in diesem konkreten Experiment. Er wirkt sich auf die gesamte Lernsequenz aus; je weniger Objekte aber zu lernen übrig bleiben, umso größer wird sein Einfluss auf das Resultat. Zusätzlich zu dem bereits genannten „Effektverwässerer“ eines weitaus nicht perfekt gelernten Merkmalswissens beeinflusst das die Trennschärfe zwischen den gewussten und den nicht-gewussten Objekten.

Abbildung 38 präsentiert eine Reanalyse des gesamten Datensatzes ohne die Nachtigall. Sie zeigt, dass die Schere bei den ObjectLosses wie anzunehmen aufgeht, und die Effektgrößen steigen allgemein an. Für drei respektive für ein nicht-gelerntes Objekt sind jetzt sogar mit Werten deutlich über 1 ausgesprochen große Effekte zu registrieren. Auch die Falsch-Alarm-Rate sinkt bei drei bzw. einer nicht-gelernten Vogelart auf akzeptablere Werte. Berücksichtigt man diese Auswertung, so kann das Verfahren also in jedem Fall bei wenigen nicht-gewussten Objekten mit hoher Effizienz eingesetzt werden.

Das Muster der ObjectLosses unterscheidet sich allerdings bei den „menschlichen“ Ähnlichkeitsurteilen etwas von der Simulation. Während in letzterer die ObjectLosses für nicht-gewusste Objekte mit erstaunlicher Konstanz um den Wert von 0.90 schwanken, steigen sie im Experiment mit der Zunahme gewusster Objekte deutlich an (von knapp 0.80 auf deutlich über 1.00). Dafür erhalten die bekannten Objekte diesmal – relativ unabhängig von der Anzahl der gelernten Vogelarten – konstante Mittelwerte um die 0.55. Das ist ein Niveau, welches in der Simulationsstudie erst bei 9 nicht-gewussten Objekten erreicht wird. Erst wenn alle Objekte gelernt sind, fällt dieser Wert im Experiment auf 0.49, während er in der Simulation dann natürlich 0.00 beträgt.

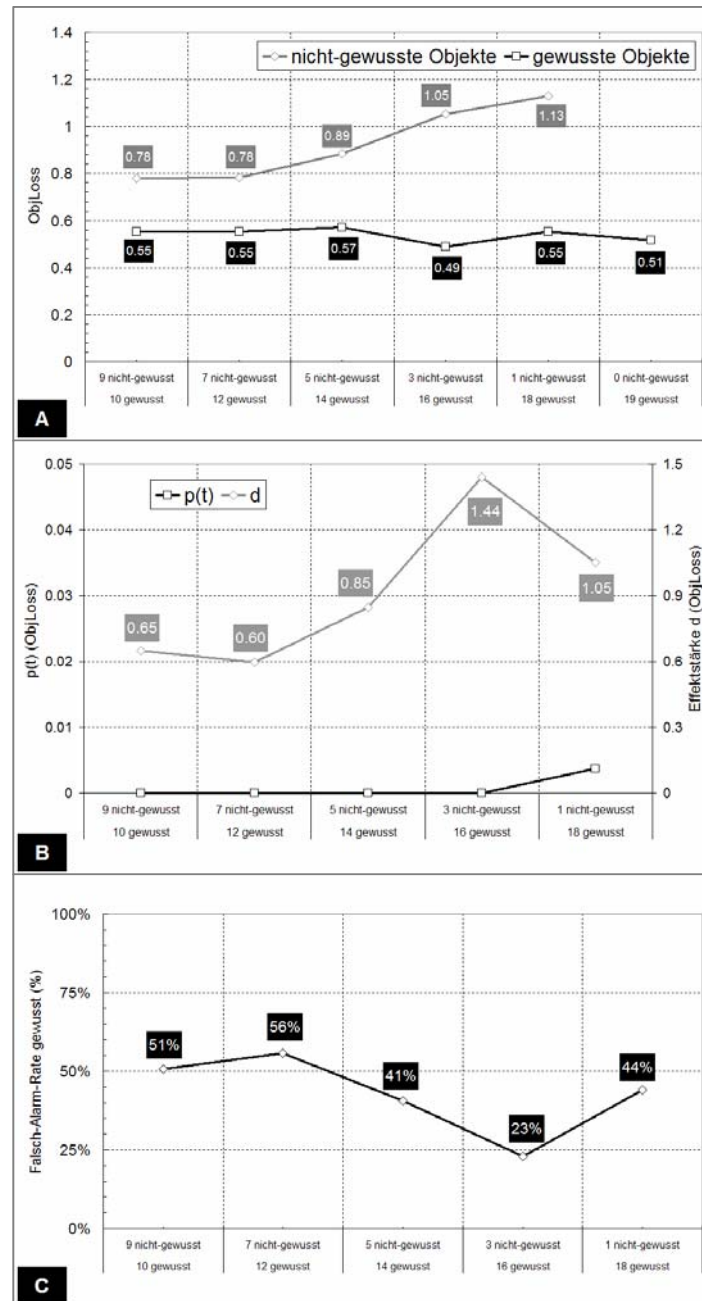


Abbildung 38. Auswertung wie in Abbildung 9, jedoch ohne das Objekt „Nachtigall“.

4.4. Gesamtdiskussion

Zusammenfassung der Resultate. Die Simulationsstudie hat die grundsätzliche Funktionsfähigkeit des vorgeschlagenen Diagnoseverfahrens aufgezeigt. Die Resultate deuten darauf hin, dass mindestens 60% der Objekte in der Wissenskarte gewusst sein sollten, damit sich durch die Größe der ObjectLosses die nicht-gewussten Objekte optimal herausfiltern lassen. Die Ergebnisse des Lernexperimentes weisen (nach Ausschaltung des „Nachtigall-Effekts“) erst ab ca. 80% gewusster Objekte eine entsprechend große Diagnoseleistung auf. Zwischen 50% und 80% gelernter Vogelarten sind die Unterschiede in den ObjectLosses zwar alle auch bei strengem alpha im t -Test signifikant, jedoch erreicht die Effektstärke mit Werten zwischen $d = .52$ und $d = .71$ nur die Region eines mittelgroßen Effektes. Hier müsste für die Anwendung

in der Praxis also im Einzelfall eine Abwägung stattfinden, ob eine etwas wacklige Diagnose ungewusster Objekte den Aufwand der Ähnlichkeitsurteile rechtfertigt.

Wie bereits erwähnt, haben die Lerneinheiten im Experiment ihren Zweck nur zum Teil erfüllt. Die Versuchspersonen verbesserten sich zwar signifikant, erreichten aber bei weitem keine Expertise hinsichtlich der gelernten Vogelarten. Durch die Schwierigkeit des hierarchisch gestuften Fragebogens war auch die Varianz zwischen gut und weniger gut gelernten Objekten zu klein, als dass man in einer Nachauswertung die Vorhersagekraft des ObjectLosses auf die „tatsächlichen“ Wissensunterschiede hätte reduzieren können. Somit bleibt also als zusammenfassende Aussage, dass auch unter den Unwägbarkeiten „menschlicher“ Ähnlichkeitsurteile der ObjectLoss zumindest dann zwischen gewussten und nicht-gewussten Objekten diskriminiert, wenn mehr als 80% der Objekte in der Wissenskarte gewusst sind. Ob und wie weit sich diese Grenze noch in Richtung der 60%-Marke aus der Simulation verschieben lässt, muss weiterer Forschung vorbehalten bleiben (eventuell in Experimenten unter noch kontrollierteren Lernbedingungen).

Es scheint uns wesentlich, noch einmal darauf hinzuweisen, dass der Terminus „gewusst“ in diesem Kontext sowohl für „korrekt gewusst“ als auch für „falsch gewusst“ steht, weil sich die beiden Formen des Wissens subjektiv für die befragte Person ja nicht unterscheiden und deswegen zu derselben Konsistenz in den Ähnlichkeitsurteilen führen. Die „magische Marke“ von 60% Vorwissen impliziert also nicht korrektes Vorwissen. Konsistent falsch platzierte Objekte werden durch die Prokrustes-Transformation in jedem Fall als solche diagnostiziert und – sofern die Abweichung beträchtlich ist – für kommende Lektionen ausgewählt. Vielleicht ist es deswegen klarer, davon zu sprechen, dass das Verfahren mit Sicherheit bis zu 20% Objekte erträgt, von denen die befragte Person keinerlei Wissen besitzt. Technisch gesehen sind sogar bis zu 40% möglich. Ohne jede Frage limitieren diese Quoten den Einsatz des Verfahrens auf Fälle, in denen man davon ausgehen kann, dass die Personen für die Mehrzahl der Objekte eine klare Vorstellung hat.

Einordnung der Befunde in die Frage der Adaptivität virtueller Lernumgebungen. Individuen, die dasselbe Sachgebiet lernen sollen, unterscheiden sich häufig hinsichtlich ihres Vorwissens massiv voneinander. Während Frontalunterricht in Klassenräumen oder Hörsälen bei der Konzeption notwendigerweise ein Level an Expertise festlegen muss, das ideal ist, um nicht unter- oder überfordert zu sein, können virtuelle Lernprogramme stärker auf den Wissensstand des einzelnen eingehen. Um in der Auswahl der Inhalte für die einzelnen Situationen hinsichtlich des individuellen Kenntnisstandes adaptiv sein zu können, benötigen virtuelle Lernumgebungen aber hinreichend gute Diagnostikmethoden. Klein (2000) diskutiert die Schwierigkeiten einer solchen Adaptivität. Er führt aus, dass Adaptierbarkeit (d.h. die Anpassbarkeit des Systems durch den Nutzer, wie z.B. die Wahl unterschiedlicher Bildschirmarrangements oder das An- und Ausschalten von Hilfefunktionen) technisch wie konzeptionell wesentlich leichter zu lösen ist als inhaltliche Adaptivität der Lerninhalte (zur begrifflichen Unterscheidung von Adaptivität und Adaptierbarkeit vgl. auch Leutner, 1992a, 1995).

In der Tat zeigt bereits ein unsystematischer Überblick über die derzeitigen virtuellen Lernprogramme, dass nur die wenigsten von ihnen im Sinne von Leutner und von Klein adaptiv

sind. Am ehesten ist dies noch für den Bereich regelbasierten Wissens der Fall, also für Physik- und Mathematikwissen (z.B. Albert & Lukas, 1999). Hier kann die Diagnose über einen vergleichsweise geringen Satz von Textaufgaben geschehen. Beim Faktenwissen hingegen bleibt der Lerner in der allermeisten Systemen auf seine eigene Einschätzung des Wissens angewiesen.

Klein (2000) hat in seinem hypermedialen Lernsystem *incops* (Introduction to Cognitive Psychology) viele seiner Vorschläge zur Adaptivität beispielhaft realisiert. Das Programm berücksichtigt seine theoretischen Annahmen hinsichtlich Interaktivität (Feedback, Fragen), Adaptivität (Generierung der Seiten und fortlaufende Wissensdiagnose: Eingangstest zur Bestimmung des Vorwissens und nach jedem Kapitel Kontrollfragen, Speicherung von Benutzerdaten für die kommende Lerneinheit), Adaptierbarkeit (Übungen, Inhaltsverzeichnis, Erscheinungsbild der Bildschirmoberfläche können an- oder abgestellt werden), Lernkontrolle (Test ob alles gelernt wurde), Lehrfunktionen (Lernhilfen wie interaktive Fragen, Einleitungen, Zusammenfassungen, Lernziele, Glossar, Literaturverzeichnis etc.) sowie Modalitäten & Codes (Bild, Text, Diagramme, Animationen). Hinsichtlich der eigentlichen Wissensdiagnostik bleibt aber noch reichlich Raum für Neuerungen: Wissen wird praktisch immer über offene Fragen (die dann im korrekten Wortlaut beantwortet sein muss) und über Multiple- oder Forced-Choice getestet.

Diesem (noch immer gültigen) Stand der technischen Möglichkeiten fügt das von uns vorgeschlagene Verfahren, die Qualität des Wissens über ein Objekt durch den ObjectLoss in einer NMDS-Karte zu messen, neue Möglichkeiten hinzu. Zum einen kann damit auf Fragen verzichtet werden, für die das System nur eine einzige korrekte Lösung akzeptiert. Zum anderen – und das halten wir für den noch wichtigeren Aspekt – misst das Ähnlichkeitsurteil nicht einfach auswendig gelernte oder noch im Kurzzeitgedächtnis befindliche Formulierungen. Es bedarf vielmehr einer gewissen Integration der gelernten Fakten, um ein relationales Urteil zu einem anderen Objekt angeben zu können. Erst wenn das vom System Angebotene auch vom Lerner wirklich verstanden ist, kann er die gestellte Aufgabe lösen. In dieser Weise misst der ObjectLoss bereits die Integration des Wissens in eine Struktur und geht damit weit über einfache Multiple-Choice-Aufgaben oder automatisiert auswertbare offene Fragen hinaus. Gleichwohl ist das Verfahren jedoch so weit automatisierbar, dass es die individuelle Zusammenstellung von Lektionen in einem virtuellen Lernsystem steuern könnte.

Einordnung der Befunde in die Verfahrensentwicklung. Es sollte nicht unerwähnt bleiben, dass mit dem hier vorgeschlagenen Einsatz NMDS-basierter Wissenskarten technisch ein Übergang von gemittelten Matrizen hin zu individuellen Matrizen vollzogen wurde. Bisherige sozial- und verhaltenswissenschaftliche Studien, die das Verfahren der Nonmetrischen Multidimensionalen Skalierung verwenden, basieren praktisch immer auf einer über eine Stichprobe gemittelten Proximitätenmatrix. Erst mit der Entwicklung eines robusten NMDS-Algorithmus wird es überhaupt möglich, stabile Karten auf individueller Ebene zu erzeugen. Außerdem werden auch die Möglichkeiten der Prokrustes-Transformation durch das vorgeschlagene Verfahren in einer neuen Weise ausgereizt. Die wenigen bislang in der Literatur vorliegenden Studien, in denen NMDS-Karten auf diese Weise übereinander gelegt wurden, interpretieren in der Regel nur die Gesamtabweichung (also den AverageLoss). Es dürfte das

erste Mal sein, dass das wesentlich detaillierte Maß des ObjectLosses im sozialwissenschaftlichen Forschungsfeld als abhängige Variable und als sogar automatisierbares Verfahren zur Auswahl zu lernender Objekte vorgeschlagen wird. In der Tat sind bislang überhaupt keine „ernsthaften“ Anwendungsergebnisse bekannt, in denen der ObjectLoss verwendet wurde (Groenen, mündliche Mitteilung).

Auch wenn die Befunde aus dem Lernexperiment die Größe des erzielbaren Prognoseeffektes noch nicht exakt abzuschätzen in der Lage sind, sollte der Nachweis, dass zumindest unter einer großen Mehrzahl gewusster Objekte die wenigen nicht-gewussten gut detektiert werden können, Anlass zu Optimismus für die weitere Forschung geben.

4.5. Literatur

- Albert, D. & Lukas, J. (Hrsg.). (1999). *Knowledge spaces. Theories, empirical research and applications*. Mahwah: Erlbaum.
- Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York: Grune & Stratton.
- Ausubel, D. P. (2000). *The acquisition and retention of knowledge: A cognitive view*. Dordrecht: Kluwer Academic Publishers.
- Bezzel, E. (1985). *Kompendium der Vögel Mitteleuropas: Passireformes, Singvögel*. Wiesbaden: Aula.
- Bonato, M. (1990). *Wissensstrukturierung mittels Struktur-Lege-Techniken. Eine graphentheoretische Analyse von Wissensnetzen*. Frankfurt/M: Lang.
- Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling – Theory and applications*. New York: Springer.
- Bortz, N. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Brünen, R., Leutner, D. & Tolxdorff, T. (1997). RADIOLIS. Ein Lehr- und Informationssystem für die Radiologie. In H. Conradi, R. Kreutz & K. Spitzer (Hrsg.), *CBT in der Medizin – Methoden, Techniken, Anwendungen - Proceedings zum Workshop in Aachen 6. – 7. Juni 1997* (S. 31-38). Aachen: Verlag der Augustinus Buchhandlung.
- Cohen, J. (1988). *Statistical power analysis for behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Der Fischer Weltalmanach (2002). *Staaten. Der FischerWeltalmanach - Online-Datenbank*. Aufgerufen 3. Juli 2002 unter http://www.weltalmanach.de/staat/staat_liste.html
- Eckert, A. (1998). *Kognition und Wissensdiagnose. Die Entwicklung und empirische Überprüfung des computergestützten wissensdiagnostischen Instrumentariums „Netzwerk-Elaborierungs-Technik (NET)“*. Lengerich: Pabst.
- Gower, J. C. & Dijksterhuis, G. B. (2004). *Procrustes problems*. Oxford: Oxford University Press.
- Hudec, K. (1992). *Vögel beobachten und bestimmen*. Hanau/Main: Werner Dausien.
- Janetzko, D. & Strube, G. (2000). Knowledge Tracking – Eine neue Methode zur Diagnose von Wissensstrukturen. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (S. 199-217). Göttingen: Hogrefe.
- Johnson, L. (1992). *Die Vögel Europas und des Mittelmeerraumes. Kosmos-Naturführer*. Stuttgart: Kosmos.
- Klauer, K. J. (1989). Allgemeine oder bereichsspezifische Transfereffekte eines Denksstrainings. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 21, 185-200.
- Klein, B. (2000). *Didaktisches Design hypermedialer Lernumgebungen. Die adaptive Lernumgebung „incops“ zur Einführung in die Kognitionspsychologie*. Marburg: Tectum. (incops-online: www.incops.de)
- Läge, D. & Daub, S. (2005). Kohonen networks and nonmetric multidimensional scaling: Matching two forms of representation. In K. Opwis & I.-K. Penner (Eds.), *Proceedings of KogWis05. The German Cognitive Science Conference 2005* (pp. 119-124). Basel: Schwabe.

- Läge, D. (2001). *Ähnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Läge, D., Daub, S., Bosia, L., Jäger, C. & Ryf, S. (2005). *Die Behandlung ausreißer-behafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag*. AKZ-Forschungsbericht Nr. 21. Zürich: Angewandte Kognitionspsychologie.
- Läge, D., Schlatter, K. & Ryf, S. (2006). *Wissensdiagnose durch sukzessive Einpassung von Objekten in eine NMDS-Karte*. Manuskript in Vorbereitung.
- Leutner, D. (1992a). *Adaptive Lehrsysteme. Instruktionspsychologische Grundlagen und experimentelle Analysen*. Weinheim: Psychologie Verlags Union.
- Leutner, D. (1992b). Das Testlängendilemma in der lernprozess-begleitenden Wissensdiagnostik. *Zeitschrift für Pädagogische Psychologie*, 24, 232-248.
- Leutner, D. (1993). Das gleitende Testfenster als Lösung des Testlängendilemmas: Eine Robustheitsstudie. *Zeitschrift für Pädagogische Psychologie*, 7, 33-45.
- Leutner, D. (1995). Adaptivität und Adaptierbarkeit multimedialer Lehr- und Informationssysteme. In L. J. Ising & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia* (S. 140-149). Weinheim: Psychologie Verlags Union.
- Leutner, D., Schulz-Brünen, B., Brünen, R., Pelikan, E., Kaminski, R., Buchner, H. & Tolxdorff, T. (1997). Computerunterstützte Lehr- und Informationssysteme in der Medizin. Instruktionspsychologische Grundlagen, Modelle und Anwendungen. In M. P. Baur, R. Fimmers & M. Blettner (Hrsg.), *Medizinische Informatik, Biometrie und Epidemiologie, Proceedings der GMD-Tagung 1996* (S. 158-162). München: Medien & Medizin Verlag.
- Mandl, H. & Fischer, F. (Hrsg.). (2000). *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*. Göttingen: Hogrefe.
- Plötzner, R., Beller, S. & Härder, J. (2000). Wissensvermittlung, tutoriell unterstützte Wissensanwendung und Wissensdiagnose mit Begriffsnetzen. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (S. 180-198). Göttingen: Hogrefe.
- Quillian, M. R. (1968). Semantic memory. In M. Minsky (Ed.), *Semantic information processing* (pp. 227-270). Cambridge, Mass.: MIT-Press.
- Scheele, B. & Groeben, N. (1984). *Die Heidelberger Struktur-Lege-Technik (SLT). Eine Dialog-Konsens-Methode zur Erhebung subjektiver Theorien mittlerer Reichweite*. Weinheim: Beltz.
- Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 39, 373-421.
- Siemer, J. & Angelides, M.-C. (1998). Towards an intelligent tutoring system architecture that supports remedial tutoring. *Artificial Intelligence Review*, 12, 469-511.
- Smith, E. E., Shoben, E. J. & Rips, L. J. (1974). Structure and process in semantic memory. A featural model for semantic decisions. *Psychological Review*, 81, 214-241.
- Streule, R., Daub, S. & Läge, D. (2005). Kohonen networks and nonmetric multidimensional scaling: Combined for individual knowledge diagnosis. In K. Opwis & I.-K. Penner (Eds.), *Proceedings of KogWis05. The German Cognitive Science Conference 2005* (pp. 189-194). Basel: Schwabe.
- United Soft Media (2001). *Der große Kosmos Vogelatlas 3.0* [CD-Rom]. München: Navigo.
- Weber, S. & Schumann, M. (2000). Concept Mapping Software Tool (COMASOTO) zur Diagnose strukturellen Wissens. In H. Mandl & F. Fischer (Hrsg.), *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken* (S. 158-179). Göttingen: Hogrefe.
- Young, F. W. (1970). Nonmetric multidimensional scaling: recovery of metric information. *Psychometrika*, 46, 357-388.

5.

**Effizienzsteigerung in der
ähnlichkeitsbasierten
Wissensdiagnostik**

Zusammenfassung⁶

Mittels Kognitiver Karten (berechnet über die Nonmetrische Multidimensionale Skalierung) können strukturelle Aspekte des objekt- und merkmalsbasierten Faktenwissens überprüft und Wissensdefizite erkannt werden. Darauf basierende adaptive Lernempfehlungen eignen sich für den Einsatz in computerbasierten, tutoriellen Systemen. Die zur Erstellung von Wissenskarten benötigte Datenstruktur bilden klassischerweise paarweise, relationale Beurteilungen zwischen einem Satz von Objekten. Die Menge ist exponentiell abhängig von der Objektzahl. Für eine praktische Anwendung muss demzufolge ein effizienteres Verfahren entwickelt werden. Die Studie überprüft anhand eines Lernexperimentes ein modular gestuftes Diagnostikkonzept. Ausgehend von einem stabilen und für das Inhaltsgebiet repräsentativ gelernten Überblickswissen ergibt sich die Möglichkeit, die benötigte Urteilsmenge von einem exponentiellen in einen linearen Zusammenhang zu wandeln, ohne dass die Kartenvalidität maßgeblich davon beeinflusst wird. Die Resultate lassen die Verwendung von Kognitiven Karten als effizientes Messinstrument für Sachwissen in virtuellen Lernsystemen lohnenswert erscheinen, nicht zuletzt auch aufgrund der gesamthaften Evaluation der individuellen Wissensstrukturen.

Schlüsselwörter: Sachwissen, Merkmalswissen, Wissenskarte, Kognitive Karte, Nonmetrische Multidimensionale Skalierung, Prokrustes-Transformation, Lernexperiment, Ähnlichkeitsurteil, Effiziente Wissensdiagnostik

5.1. Einleitung

Wissensdiagnostik. Das deklarative Merkmalswissen einer Person über eine Anzahl von Objekten lässt sich mit so genannten „Kognitiven Karten“ darstellen (Läge, 2001, Marx & Hejj, 1989; Marx & Läge, 1995). Als dimensionale Gebilde zeigen sie die Relationen, die eine Person zwischen diesen Objekten sieht, in Form von kleineren oder größeren Distanzen auf und lassen als Gesamtbild damit Rückschlüsse auf die von der Person als wesentlich erachteten Merkmalsbereiche zu. Gleichzeitig erlauben Kognitive Karten, das Wissen über ein spezifisches Objekt durch deren Position in der Gesamtstruktur nachzuvollziehen und die Wissensqualität zu bewerten: Wir konnten zeigen, dass sich Lernfortschritte auf der Gesamtstrukturebene nachzeichnen lassen (vgl. Kapitel 2 in dieser Arbeit). Kognitive Karten erlauben ebenfalls, kognitive Äquilibrationsprozesse beim Lernen wie beispielsweise Assimilation und Akkomodation zu messen (vgl. Kapitel 3 in dieser Arbeit). Es existieren sogar starke Hinweise darauf, dass unter bestimmten Voraussetzungen Vorhersagen über die Wissensqualität auf Einzelobjektebene gemacht werden können (vgl. Kapitel 4 in dieser Arbeit).

⁶ Teile dieser Arbeit wurden an der 46. Tagung experimentell arbeitender Psychologen (TeaP) 2004 in Giessen als Poster vorgestellt (Streule, R., Egloff, S. & Läge, D. (2004). Effizienz in der ähnlichkeitsbasierten Wissensdiagnostik. In D. Kerzel, V. Franz & K. Gegenfurtner (Hrsg.), *Beiträge zur 46. Tagung experimentell arbeitender Psychologen. Justus-Liebig-Universität Giessen. 04. April – 07. April 2004*. Lengerich: Pabst. S. 259.).

Abbildung 39 gibt ein Beispiel für eine solche Wissenskarte. Sie wurde mittels *Nonmetrischer Multidimensionaler Skalierung (NMDS)* aus einer Matrix paarweiser Ähnlichkeitswerte generiert (zur Logik des Verfahrens vgl. Borg & Groenen, 1997). In diesem Fall handelt es sich um die aus den Merkmalsdaten des hier vorgestellten Experimentes errechnete Normkarte, aber man kann sich als Einführungsbeispiel auch vorstellen, dass diese NMDS-Karte auf der Basis eines Satzes individueller Ähnlichkeitsurteile entstanden ist. Durch Property Fitting (Läge, 2001) ließen sich im Nachhinein über multiple Regressionsanalysen die drei Dimensionen „Bruttosozialprodukt“, „Landesfläche“ und „Einwohnerzahl“ mit genügend guten Korrelation in die Karte einpassen, so dass im Fall einer individuellen Wissenskarte interpretiert werden dürfte, dass die befragte Person bei ihren Urteilen vor allem auf diese Merkmale zurückgegriffen haben muss. Außerdem lässt sich bereits aus der Relation zu den anderen Objekten für ein beliebiges Land, z.B. für Deutschland, sagen, dass die Person hierunter ein relativ großflächiges, einwohnerreiches Land mit hohem Bruttosozialprodukt versteht.

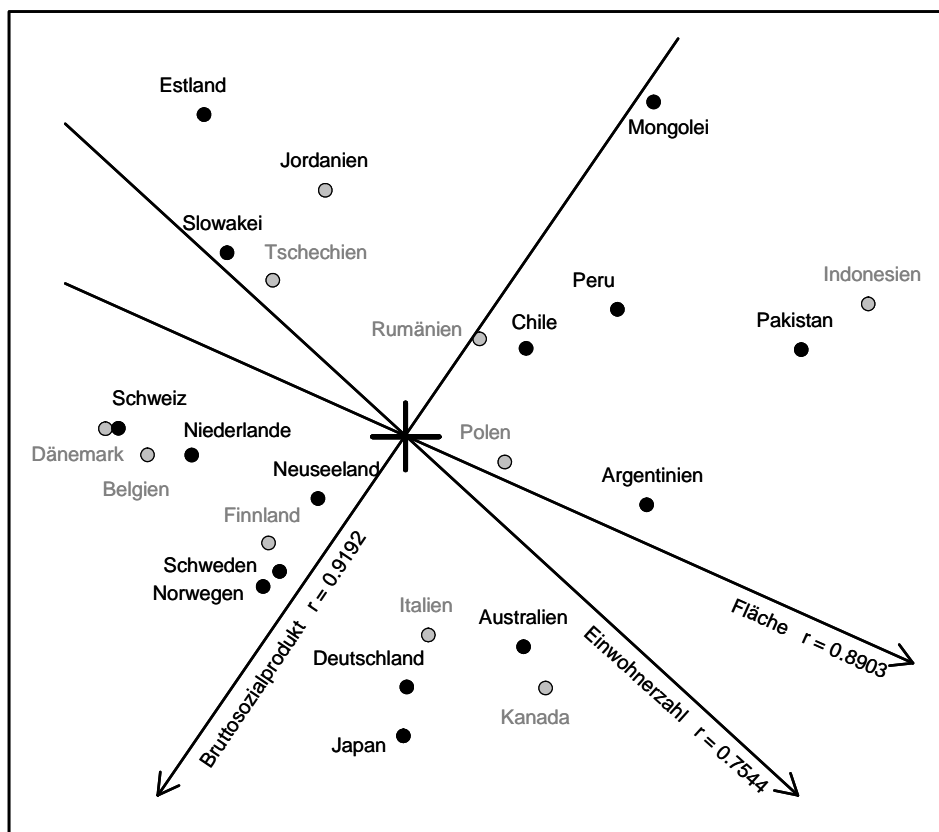


Abbildung 39. NMDS-Lösung bestehend aus 25 Objekten, welche in diesem Experiment als Normkarte Verwendung findet (Schwarze Punkte stehen für die 15 Objekte der Grundstruktur, graue für die erweiterte Vollstruktur, bestehend aus 25 Objekten: vgl. dazu Methode - Normmodell).

Solch individuelle „Landkarten“ können durch *Prokrustes-Transformation* (Gower & Dijksterhuis, 2004) mit einem Expertenmodell verglichen werden. Bereits ohne großen mathematischen Aufwand springt bei diesem Vergleich ins Auge, welche Objekte ein Lerner gut kennt (d.h. korrekt verortet) und welche er falsch platziert (als Beispiel vgl. Abbildung 40). Numerisch lässt sich diese Abstandsinformation zweier korrespondierender Objekte zwischen

Ist und Soll als „Object Loss“ (kurz: ObjLoss) ausdrücken. Die Gesamtabweichung wird als „Average Loss“ (kurz: AvgLoss) quantifiziert (Läge, 2001).

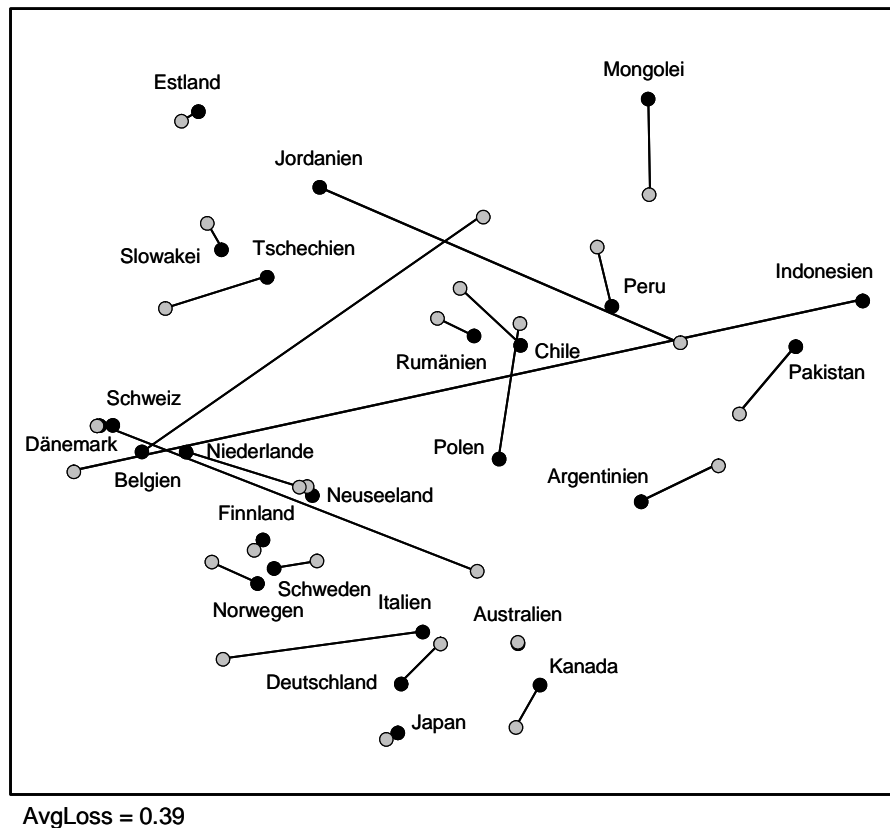


Abbildung 40. Ergebnis einer Prokrustes-Transformation. Die Objektpositionen des Normmodells sind schwarz, diejenigen der (fiktiven) Lernerkarte grau gekennzeichnet. Die mittlere Differenz aller Objektabweichungen (= ObjLoss; durch Striche verbunden) ergibt die Gesamtabweichung AvgLoss = 0.39.

Problemstellung. Grosse Abweichungen im ObjLoss deuten auf mangelhaftes Merkmalswissen hin. In der praktischen Anwendung dieser Verfahrenslogik in einem Lehr- / Lernsystem bietet es sich an, anhand dieser Performanzschätzung den Lerninhalt für kommende Lektionen zusammenzustellen. Intelligente Tutorielle Systeme (ITS) oder Adaptive Tutorielle Systeme (ATS) – als Beispiele für virtuelle Lernprogramme – verfügen über Lehrstoffexpertise und erlauben dadurch inhaltliche Adaptivität, indem sie sich auf der instruktionalistischen Seite an den Wissensstand des Lerners anpassen (vgl. zu den Komponenten eines ITS Van Merriënboer, Van Den Berg & Maaswinkel, 1989 und Leutner, 1992a). Die Verwendung der NMDS und Prokrustes-Transformation als Diagnosemodule eröffnet zudem die Möglichkeit, das gesamte Wissen auf strukturell-relationaler Ebene zu messen (als Vorteil im Vergleich zu einer stichprobenartigen Wissensüberprüfung in einem einfach gestrickten Multiple-Choice-Test beispielsweise). Der Differenziertheit, Komplexität und Vollständigkeit der Gedächtnisinhalte wird somit Rechnung getragen und der Diagnostikprozess orientiert sich stärker an theoretischen Modellen der Gedächtnispsychologie (so genannte Netzwerktheorien). ITS-Systeme bedingen einen gestuften Lernprozess über sich abwechselnde Lern- und Testphasen. Immer wieder wird der momentane Wissensstand des Lerners überprüft, um maximal adaptiv

sein zu können (Leutner, 1992a). Daraus folgt eine wichtige Schwierigkeit, auf welche bereits Leutner (1992b, 1993) hingewiesen hat. Das von ihm bezeichnete „Testlängendilemma“ verlangsamt den Lernprozess, wenn der Lerner mehrmals sein Wissen kundtun muss, denn die Zeit zur Diagnose geht logischerweise von der potentiellen Lernzeit ab. Dieses Problem verschärft sich bei der Verwendung der NMDS als Diagnoseinstrument zusätzlich.

Um die für die Erstellung einer Kognitiven Wissenskarte notwendige Datengrundlage zu erhalten, müssen in der klassischen Form alle paarweisen Ähnlichkeitsurteile zwischen den interessierenden Objekten (n) für eine 2-Way-1-Mode Matrix erhoben werden (Die Bestimmung der Urteilsmenge ergibt sich über die Formel: $n * (n - 1) / 2$). Die Anzahl an solchen Urteilen ist somit exponentiell von der Objektzahl abhängig. Bei 20 Objekten müssen beispielsweise 190, bei 40 Objekten bereits 780 (!) Ähnlichkeiten beurteilt werden. Nur vereinzelt wird in der empirischen Forschung solch ein großer Aufwand betrieben, wie das beispielsweise Lee, Byatt & Rhodes (2000) mit ihren Probanden für die große Urteilsmenge von 1'176 Ähnlichkeitseinschätzungen bei 49 Objekten zur Beurteilung von Gesichtern durchgeführt haben. Für experimentelle Laborstudien mit einer klar formulierten Fragestellung mag dieser Aufwand gerechtfertigt sein, für die Praxisanwendung in computerbasierten Lernumgebungen ist dies jedoch ein nicht durchführbares Unterfangen. Das anzustrebende Ziel in der ähnlichkeitsbasierten Wissensdiagnostik muss folgerichtig eine effiziente, aber dennoch valide Erhebungsstrategie sein, um im Sinne von Leutner (1992a, 1992b) maximal adaptiv und trotzdem so ökonomisch wie möglich zu sein. Doch wie könnte eine Effizienzsteigerung erreicht werden?

Nonmetrische Multidimensionale Skalierungsverfahren fokussieren auf ordinale Verhältnisse. Im Gegensatz zu einer klassischen Multidimensionalen Skalierung (MDS), welche metrische Beziehungen skaliert, besitzt die nonmetrische Variante mehr Freiheitsgrade in der räumlichen Anordnung der Objekte, da für eine korrekte Darstellung der Rohdaten lediglich Rangreihenfolgen eingehalten werden müssen. NMDS-Verfahren beinhalten approximative Iterationsalgorithmen, welche, je nach Methode, unterschiedliche Parameter zu minimieren versuchen. Bei unvollständiger Datenbasis beispielsweise erhalten die einzelnen Proximitäten (= relationale Daten) ein viel höheres Gewicht bei der räumlichen Skalierung. Daraus folgt, dass eine NMDS-Lösung umso stabiler wird und die Matrixwerte und ergo auch die strukturellen Charakteristika adäquater und eindeutiger abbilden kann, je vollständiger sich die Datengrundlage zeigt. Der einfachste Weg, dem Problem einer instabilen Kognitiven Karte aus dem Weg zu gehen, ist, dass ausschließlich vollständige, asymmetrische Dreiecksmatrizen skaliert werden, wie dies im Fall des MINISSA-Algorithmus (Roskam & Lingoes, 1970) umgesetzt wurde. Effizienter wäre der Diagnostikaufwand aber unbestritten, wenn die Kartenberechnung auch mit fehlenden Matrixwerten („missing values“) umgehen könnte. Fraglich bleibt, welches das forschungsbezogene beste Vorgehen und nicht aus rein mathematischen Überlegungen optimalste Skalierungsverfahren ist. Ein Vergleich von einigen anerkannten NMDS-Verfahren soll diesen Punkt konkretisieren.

Wie bereits erwähnt, kann MINISSA mit unvollständigen Matrizen nicht umgehen. Der ALSCAL-Algorithmus (Takane, Young & de Leeuw, 1977; ausführlich beschrieben bei Young & Lewyckyj, 1979) demgegenüber benötigt ebenfalls eine vollständige Ausgangsmat-

rix. Bei fehlenden Werten wird als Schätzwert einfach der Mittelwert aller vorhandenen Matrixwerte verwendet. Das Ziel von ALSCAL ist es, die Punktekongfiguration als Ganze bestmöglich abzubilden. Der Nachteil dieses Konzepts ist aber, dass damit eine Einengung in der Differenzierung von einzelnen Objektpositionierungen riskiert wird, da zwar die Mittelwertbildung nicht zwangsläufig eine Verschlechterung der Urteilkonsistenz bedingt, jedoch die Wahrscheinlichkeit steigt, dass die betreffenden Objekte eher in eine mittlere Region der Karte zu liegen kommen aufgrund der sich in einer mittleren Stärke befindlichen Ähnlichkeitseinschätzungen zu allen anderen Objekten. Die optimistische Annahme, dass die schlussendliche Positionierung der „wahren“ entspricht, muss sehr in Frage gestellt werden. Der PROXSCAL-Algorithmus (Commandeur & Heiser, 1993), einer der am meisten verwendeten NMDS-Verfahren⁷, verzichtet auf eine nachträgliche Schätzung fehlender Daten und schließt diese leeren Zellen einfach aus der Berechnung aus. Dasselbe Vorgehen verfolgt auch der in dieser Studie verwendete RobuScal-Algorithmus (Läge, Daub, Bosia, Jäger & Ryf, 2005), allerdings unterscheiden sich diese beiden Verfahren in einem anderen zentralen Punkt, nämlich in der Robustheit. Die Verwendung von quadrierten Fehlermodellen bei der Generierung der optimalen Punktekongfiguration (wie bei MINISSA, ALSCAL und PROXSCAL) haben den gewichtigen Nachteil, dass Ausreißer in den Daten überproportional stark in den Skalierungsvorgang einfließen. Auf einer inhaltlichen Ebene sind es aber nicht die Ausreißer (Inkonsistenzen, Transitivitätsverletzungen) in den Ähnlichkeitsurteilen, welche interessieren, sondern die einzelnen Objektpositionen (im Gegensatz zum erwähnten Ziel von ALSCAL bspw.). Erst ein Gewichtungsverfahren, welches unpassende Urteile weniger stark berücksichtigt und somit das „Signal“ vom „Rauschen“ trennt, erlaubt die für Wissensabschätzungen notwendige Interpretation von Kognitiven Karten auf individueller, personenspezifischer Ebene. RobuScal besitzt ein Gewichtungsmodul, indem er Inkonsistenzen detektiert und in der Positionierung, je nach Stärke der Inkonsistenz, Urteile weniger oder mehr runtergewichtet. Die Gewichtung fehlender Werte wird bei RobuScal auf Null gesetzt, womit die betreffenden Objekte nur einen geringen Einfluss auf die räumliche Darstellung der anderen Objekte haben. Bei der am Schluss durchgeführten Abschätzung der Abbildungsgüte (anhand des NMDS-Stress; vgl. Läge et al., 2005 und Borg & Groenen, 1997) werden diese Distanzen zwischen den Objekten ebenfalls nicht einbezogen.

Auch wenn RobuScal analog der geführten Argumentation psychologisch betrachtet mit unvollständigen Daten umgehen können sollte, bleiben Unklarheiten über deren Systematik bestehen. Die einfachste Form wäre ein zufälliges im Gegensatz zu einem systematisch strukturierten Lückensetzen. Ein zufälliges Weglassen von Urteilen beeinflusst aber die Robustheit des Skalierungsverfahrens. In simulativen und experimentellen Studien konnte gezeigt werden, dass der Matrixumfang bei der Verwendung von RobuScal für eine adäquate räumliche Darstellungsgüte der Ähnlichkeitsurteile mindestens 15 (Läge et al., 2005), besser aber 20 Objekte (vgl. Kapitel 4 in dieser Arbeit) umfassen sollte. Werden nun in diese minimal notwendige Urteilsmenge zusätzlich noch Lücken gesetzt, kann nicht mehr von einer robusten

⁷ Proxscal ist neben dem ALSCAL-Algorithmus im Statistikpaket SPSS, einem weit verbreiteten Analyseprogramm, implementiert.

Objektpositionierung ausgegangen werden und somit würden zu viele Fehleinschätzungen bezüglich des Objektwissens entstehen. Der Weg zu mehr Effizienz führt dementsprechend nicht an einer Vollerhebung von mindestens 15 Objekten vorbei und es entfällt die Möglichkeit eines zufälligen Lückensetzens.

Lösungsansatz. Das Erkennen von Wissensdefiziten anhand NMDS und Prokrustes-Transformation funktioniert nur unter bestimmten Voraussetzungen. Eine genügend hohe Vorhersagegenauigkeit bezüglich des Lernerwissens bedingt ein bestimmtes Maß an Vorwissen. Erst wenn eine Person über eine bestimmte Anzahl an Objekten stabile Repräsentationen besitzt, greift die Robustheit des Skalierungsverfahrens und mangelhaft gespeicherte Objekte werden überzufällig detektiert (vgl. Kapitel 4 in dieser Arbeit). Dieses Faktum, kombiniert mit der vorher erwähnten technischen Einschränkung einer Mindestzahl von 15 Objekten, erlaubt den Einsatz der NMDS zur Wissensmessung erst, wenn ein bestimmtes Grundwissen vorhanden ist. Außerdem muss, wie gesehen, dazu die vollständige Datenmatrix erhoben werden (Beispiel: Bei 15 Objekten ergeben sich 105 Ähnlichkeitsurteile. Die benötigte Bearbeitungszeit bewegt sich erfahrungsgemäß bei ca. 10 bis 15 Minuten).

Der *erste Effizienzgewinn* ergibt sich bei Fortschreiten des Lernvorgangs und der Erweiterung des Wissens. Ein stabiles Vorwissen (Grundstruktur) erlaubt die adäquate Darstellung der räumlichen Objektverhältnisse. Die stabilisierenden Kräfte dieser Grundstrukturobjekte eröffnen die Perspektive, dass bei der Diagnose neu gelernten Wissens nur noch die Ähnlichkeitseinschätzungen zwischen den neuen und den Objekten der Grundstruktur erhoben werden müssten. Es könnte also auf die Urteile innerhalb der Basisobjekte und innerhalb der neuen Objekte verzichtet werden, was beispielsweise bei 15 Grundstruktur-Objekten und 10 neuen Objekten eine Verringerung des Diagnostikaufwandes von 50% bedeutet (anstatt 300 nur noch $10 \text{ mal } 15 = 150$ Urteile). Je mehr neues Wissen gelernt wird, desto höher wird der Effizienzgewinn, da sich der ehemals exponentielle Zusammenhang zwischen Objekt- und Urteilsmenge in einen linearen gewandelt hat. Eine *zweite Effizienzsteigerung* ergibt sich dann über den Verzicht an Daten innerhalb der Urteile zwischen neu gelernten Objekten und den Grundstrukturobjekten. Es werden mindestens Urteile zu drei anderen Objekten benötigt, um einen Punkt in einem euklidischen Raum eindeutig zu positionieren. Offen bleibt, wie nah an dieses geometrische Minimum reduziert werden kann, ohne dass die Abbildungsgüte leidet.

Operationalisierung. Die Durchführung von Effizienzgewinn 1 ist nur möglich, wenn von einem über die Zeit stabilen Vorwissen ausgegangen werden kann. Erst die strukturbildenden Kräfte der Grundstrukturobjekte erlauben eine exakte Urteilsskalierung und Abschätzung der Wissensqualität von neu Gelerntem. Aus gedächtnistheoretischer Sicht sollte dies kein Problem darstellen, da Wissen, einmal im Langzeitgedächtnis gespeichert, über die Zeit nur langsam verwischt (mal abgesehen von kognitiven Gedächtnistäuschungen wie etwa dem Hindsight-Bias). Zu zeigen gilt es aber, ob diese hypothetische Annahme auch bei der Verwendung der NMDS als Messinstrument gilt, dass sich also Kognitive Karten bei Messwiederholungen und gleichem Wissenstand nicht maßgeblich voneinander unterscheiden. Bestimmte Unterschiede (operationalisiert über den AvgLoss einer Prokrustes-Transformation) sind zu erwarten, da immer noch gewisse Unterschiede im Urteilsverhalten auftreten können. Als Kriterium für die tolerierbare Abweichung wird die *intrapersonale Kartenidentität* herange-

zogen. Für den Zweck der adaptiven Wissensdiagnostik muss die Ordnung der ObjLosses der individuellen Karten bei Messwiederholung im Vergleich zur Normkarte innerhalb einer Person gleich bleibend sein (*kognitive Stabilität* bzw. Reliabilität des Verfahrens). Ist dies gegeben, werden sich die Karten derselben Person signifikant weniger voneinander unterscheiden als die Karten zwischen verschiedenen Probanden ($\text{AvgLoss}_{\text{intrapersonal}} < \text{AvgLoss}_{\text{interpersonal}}$).

Des Weiteren gilt zu zeigen, dass bei der zusätzlichen Skalierung neuer Objekte die Positionen der Grundstrukturobjekte vergleichbar bleiben (*methodische Stabilität*). Nur wenn dies gegeben ist, kann eine einmal gelernte Grundstruktur als gegeben betrachtet werden, wenn neues Wissen für sich alleine skaliert werden wird. Auch hier ist ein gewisser Unterschied zu erwarten, da die NMDS eine größere Anzahl an relationalen Beziehungen skalieren muss⁸. Aufgrund des iterativen Approximationsalgorithmus wird die bestmögliche Abbildungsgüte der Matrixwerte gewisse Objektverschiebungen mit sich bringen. Zwei Abbruchkriterien können angesetzt werden: Das strengere Kriterium ergibt sich über den Vergleich der Positionen der Grundstrukturobjekte, wenn sie zum einen per se und zum anderen inklusive der neuen Objekte skaliert werden. Die Normkarte basiert auf mathematisch exakt bestimmten relationalen Verhältnissen zwischen den Objekten und besteht aus zueinander perfekt konsistenten Urteilen. Die Abweichung, welche sich aufgrund der unterschiedlichen Objektmenge ergibt, bildet den Richtwert für die zu erwartende Positionsverschiebung. Dieser Wert ist ausschließlich auf das Skalierungsverfahren zurückzuführen und dessen Umgang mit der auf gleicher Dimensionalität, aber unterschiedlicher Datenmenge basierender Prozedurcharakteristik. Als zweites und etwas schwächeres, empirisches Ablehnungskriterium wird ein Kartenabweichungswert von $\text{AvgLoss} = 0.28$ angesetzt. Dieser Wert ergibt sich aus der Befragung von Ornithologie-Experten, welche zum einen Ähnlichkeitsurteile zu 20 Objekten und zum anderen Urteile zu lediglich 10 aus denselben 20 Objekten lieferten. Der Vergleich der Koordinaten der korrespondierenden 10 Objekte aus den zwei Erhebungen ergab den vorher genannten Wert (unpublizierte Daten). Die Relevanz dieses Maßes ergibt sich durch die Annahme, dass bei Experten die Qualität des Wissens zwischen den Erhebungszeitpunkten nur marginalen Veränderungen unterliegt und dementsprechend die Kartenunterschiede ein reines, nicht kontrollierbares Rauschen im Urteilsverhalten darstellen (aufgrund der kognitiv gleich bleibenden ordinalen Verhältnisse zwischen den Objekten, ungeachtet dessen, dass zum einen nur 10 und zum anderen 20 Objekte beurteilt wurden) und nicht auf das Skalierungsverfahren zurückzuführen sind. Das Inhaltsgebiet Ornithologie ist zudem komplexer strukturiert als das in dieser Studie verwendete des Länderwissens (vgl. Abbildung 39), was bedeutet, dass der Wert von 0.28 als obere Grenze zu verstehen ist.

Der Ablehnungsbereich bei Anwendung von Effizienzgewinn 1 ergibt sich wie bereits bei der zeitlichen Stabilitätsüberprüfung anhand der Beibehaltung der individuellen Kartenidentität.

⁸ Postuliert werden könnte auch eine kognitive Umstrukturierung im Sinne einer Akkomodation (Piaget, 1976) bei Neuaufnahme von Wissen. Über einen kontrollierten didaktischen Lernprozess kann aber von reinen assimilativen Integrierungen ausgegangen werden, da sich die dimensionale Struktur des Vorwissen nicht ändern muss, wenn lediglich zusätzliches, aber zu den bestehenden Repräsentationen kongruentes Lernmaterial aufgenommen wird (Ausubel, 2000).

Werden nur noch Relationen zwischen den neuen Objekten und jenen der Grundstruktur erhoben, muss gezeigt werden, dass im Vergleich zu einer Kartenberechnung mit allen Daten die ordinale individuelle ObjLoss-Ordnung beibehalten wird. Sind die intrapersonalen Kartenabweichungen zwischen den zwei Berechnungsverfahren kleiner als die interpersonalen, kann davon ausgegangen werden. Außerdem dürfen die absoluten Werte nicht höher sein als jene, welche sich über Messwiederholungen ergeben, da nur dann ein Unterschied auch anhand des Erhebungszeitpunktes erklärt werden könnte.

Der Verzicht auf Urteile innerhalb der Vergleiche zwischen den neu gelernten Objekten und jenen der Grundstruktur versucht, die Ökonomie in der Datenerhebung noch zu vergrößern. Sukzessive Reduktionsstufen loten die Schwelle aus, ab derer der Effizienzgewinn 2 zu einer zu großen Verzerrung der Abbildungsgüte führt. Verglichen werden die Objektpositionen der neuen Objekte, basierend auf einer Vollerhebung, mit den Objektpositionen, welche sich aufgrund der einzelnen Reduktionsstufen ergeben. Die mittlere Kartenabweichung zwischen zwei Erhebungszeitpunkten bildet das Abbruchkriterium. Liegen die Objektverschiebungen (gemessen anhand der gemittelten ObjLoss-Werte der neuen Objekte) unter dem Wert aus Messwiederholungen, kann nicht mehr eindeutig geschlossen werden, ob die Unterschiede auf mangelnde Skalierungsqualität oder auf den Testzeitpunkt zurückzuführen sind.

5.2. Methode

Stichprobe. 16 Versuchspersonen (4 Frauen, 12 Männer) im Alter zwischen 25 und 29 Jahren bildeten die Stichprobe des Lernexperiments. Alle Probanden sind Studierende der Universität Zürich oder verfügen über eine höhere Berufsausbildung.

Lernmaterial. Das zu lernende Wissensgebiet der Geographie umfasst 25 Länder mit den drei Merkmalen Einwohnerzahl, Fläche in km² und Bruttosozialprodukt pro Kopf in US\$. Zur Grundstruktur zählen 15 repräsentativ ausgewählte Länder (Australien, Argentinien, Chile, Deutschland, Estland, Japan, Mongolei, Neuseeland, Niederlande, Norwegen, Pakistan, Peru, Schweden, Schweiz, Slowenien). Bei der Auswahl dieser Länder wurde darauf geachtet, dass sie auf den drei Merkmalsdimensionen möglichst unterschiedliche Ausprägungskombinationen aufweisen und somit den kognitiven Raum möglichst gut „aufspannen“ (= eine didaktisch optimale Ankerstruktur bilden). Die restlichen 10 Länder bilden das Material zur Erweiterung des Grundwissens (Belgien, Dänemark, Finnland, Indonesien, Italien, Jordanien, Kanada, Polen, Rumänien, Tschechien). Die Versuchspersonen wurden instruiert, während des Experimentes anderes Wissen über diese Länder nicht zu berücksichtigen und nur aufgrund der drei Merkmale zu entscheiden. Auf Lernkärtchen war pro Land die entsprechende Merkmalsinformation aufgeführt. Die Daten stammen aus dem Fischer Weltalmanach (2002).

Normmodell. Auf der Basis der realen Daten konnten rechnerisch Unähnlichkeiten zwischen den einzelnen Ländern bestimmt werden (vgl. die Formel in Abbildung 41); der Kehrwert stellt die Ähnlichkeit dar.

$$U\ddot{A}_{x,y} = \frac{|(x_i - y_i)/(x_i + y_i)| + |(x_j - y_j)/(x_j + y_j)| + |(x_k - y_k)/(x_k + y_k)|}{3}$$

Abbildung 41. $U\ddot{A}_{x,y}$: Unähnlichkeitsquotient normiert zwischen 0 (minimale Unähnlichkeit / maximale Ähnlichkeit) und 1 (maximale Unähnlichkeit / minimale Ähnlichkeit); x, y stehen für die Länderindizes; die Indizes i, j, k für die drei Merkmalsdimensionen

Aus diesen Proximitäten wurde mit RobuScal (Läge, Daub, Bosia, Jäger & Ryf, 2005) die in Abbildung 39 ersichtliche zweidimensionale Normkarte, bestehend aus 25 Objekten erstellt. Die Grundstruktur mit 15 Objekten berechnet sich in analoger Weise.

Versuchsdesign. In Abbildung 42 ist schematisch der Experimentalablauf dargestellt.

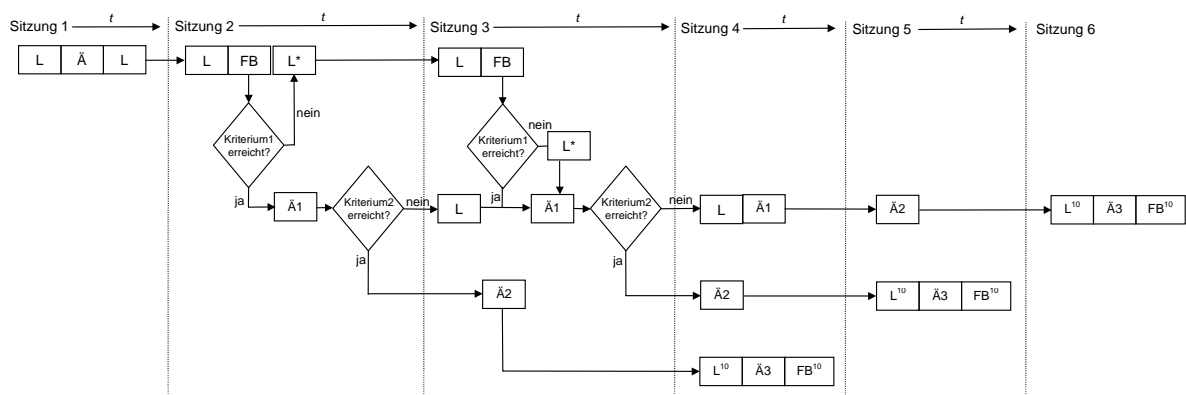


Abbildung 42. Versuchsdesign; Nähere Erläuterungen im Text.

Das Ziel des Kriteriumslernens in den ersten zwei bis drei Sitzungen war es, das Wissen der Probanden über die Grundstruktur möglichst nahe an die berechnete normative Grundstruktur heranzuführen. Jede Sitzung begann mit einer *Lernphase* (L) von 15 Minuten. Die Versuchspersonen lernten die 45 Merkmalsinformationen (15 Länder mal 3 Merkmale) selbstgesteuert, sie konnten das Lernmaterial ordnen oder sich Notizen machen. In der ersten Sitzung wurden sie zwischen den Lernphasen mit der Methodik der *Ähnlichkeitsurteile* (Ä) vertraut gemacht. In den nachfolgenden Sitzungen fand zu diesem Zeitpunkt die Überprüfung der Kriterien statt. Die einzelnen Lernsitungen fanden in *zeitlichen Abständen* (t) von zwei bis drei Tagen statt.

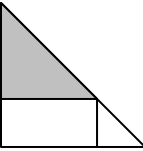
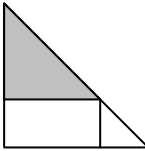
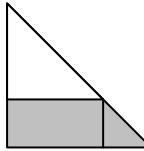
Um von einem stabilen Wissen bezüglich der drei gelernten Merkmale ausgehen zu können, sollten die Versuchspersonen über ein gutes Faktenwissen und ein hinreichendes strukturelles Wissen verfügen. Dazu mussten die Versuchspersonen zwei verschiedene Kriterien erfüllen oder mindestens drei Lernsitungen absolvieren. Mit einem *Multiple-Choice-Fragebogen* (FB) über alle 15 Länder wurde das Faktenwissen getestet (insgesamt 45 Aufgaben à 5 Antwortalternativen). Um dieses Kriterium 1 zu erfüllen, mussten von maximal 90 Punkten (30 Länder mal 3 Merkmale mal max. 2 Punkte) mindestens 75 und mindestens 5 von 6 möglichen Punkten pro Land erreicht werden (3 Merkmale mal max. 2 Punkte; korrekte Antwort = 2 Punkte; relational richtige Antwort = 1 Punkt; falsche Antwort = 0 Punkte). Wer dieses Kri-

terium 1 in der dritten Lernsitzung knapp nicht erreichte, wurde nach einer zusätzlichen *Lernphase der Wissenslücken* (L^*) à ca. 5 bis 10 Minuten trotzdem zu den Ähnlichkeitsurteilen ($\ddot{A}1$) zugelassen. Alle 15 Länder mussten anhand paarweiser *Ähnlichkeitsurteile* ($\ddot{A}1$) auf einer Skala von 1 (minimale Ähnlichkeit) bis 9 (maximale Ähnlichkeit) miteinander verglichen werden. Die aus diesen Urteilen erstellten Kognitiven Karten wurden mit der berechneten Normstruktur verglichen (Prokrustes-Transformation). Ein AvgLoss von 0.3 wurde als Abweichung toleriert, um auch das Kriterium 2 der strukturellen Übereinstimmung zu erfüllen. Zur Bestimmung dieses Kriteriums wurden die Ergebnisse eines vergleichbaren Lernexperiments (15 Länder, Merkmale: BSP, Fläche, Einwohnerzahl) herangezogen. Nach vier Lernsitzungen erreichten die Versuchspersonen im Mittel eine Abweichung von der Normstruktur von AvgLoss = 0.3 (vgl. Kapitel 2 in dieser Arbeit). Wer dieses Kriterium 2 in der vierten Sitzung nur marginal überschritt, konnte das Experiment trotzdem fortsetzen. Nach Beendigung des Kriteriumslearnens erfolgte nach einer Latenzphase von 2 bis 3 Tagen ohne zusätzliches Lernen die zweite Beurteilung der Grundstruktureobjekte ($\ddot{A}2$). Der Vergleich der Wissenskarten aus $\ddot{A}1$ mit $\ddot{A}2$ über die Prokrustes transformation bildet die Datengrundlage zur Überprüfung der zeitlichen Stabilität.

In der *letzten Lernsitzung* (L^{10}) hatten die Versuchspersonen während 20 Minuten 10 weitere Länder anhand der drei Merkmale Bruttosozialprodukt, Einwohnerzahl und Fläche in km^2 zu lernen. Anschließend beurteilten die Probanden die 10 neuen Länder nur noch mit den 15 Ländern der Grundstruktur, sowie die neuen Länder untereinander auf ihre Ähnlichkeit hin ($\ddot{A}3$). Die Kontrolle des Faktenwissens zu den neuen Ländern übernahm ein weiterer Fragebogen (FB^{10}).

In Tabelle 2 ist zur Veranschaulichung der einzelnen Erhebungszeitpunkte ein Überblick über die erhobenen Rohdaten dargestellt.

Tabelle 2. Überblick über die erhobenen Rohdaten. Die Dreiecke symbolisieren die Matrizen mit Ähnlichkeitsurteilen. Grau unterlegt sind die gemessenen Urteile.

$\ddot{A}1$ (Ähnlichkeitsurteile)	$\ddot{A}2$ (Ähnlichkeitsurteile)	$\ddot{A}3$ (Ähnlichkeitsurteile)	FB (Fragebogen)	FB^{10} (Fragebogen)
105 Urteile zwischen allen 15 Ländern der Grundstruktur	105 Urteile zwischen allen 15 Ländern der Grundstruktur	150 Urteile zwischen den 10 neuen Ländern und der Grundstruktur 45 Urteile zwischen allen 10 neuen Ländern.	45 Fragen (jeweils 3 Merkmale für 15 Länder)	30 Fragen (jeweils 3 Merkmale für 10 Länder)
				
Kriterium (Fakten- und Strukturwissen)	Zeitliche Stabilität; Konsistenz	Neue Objekte	Kriterium (Faktenwissen)	Kontrolle (Faktenwissen)
	$\ddot{A}2$ und $\ddot{A}3$ addieren sich zur Vollmatrix (25 Länder) zur Untersuchung der Effizienzsteigerung			

Berechnung der NMDS-Karten. Als Kontrollbedingung wird die *Vollmatrix* (s. Tabelle 2) herangezogen. Diese besteht aus allen paarweisen Ähnlichkeitsurteilen. Zur Überprüfung der methodischen Stabilität und der Urteilreduzierung wurde ein gestuftes Positionierungsverfahren gewählt. Bei dieser *sequentiellen Teilmatrix* werden die Ähnlichkeitsurteile zwischen den neuen Objekten nicht berücksichtigt. Jedes neue Objekt wird einzeln in die ursprüngliche Grundstrukturkarte eingefügt. Für die Positionen der neuen Objekte werden diese einzeln an die Grundstruktur-Matrix angehängt und daraus eine Karte berechnet. Die so entstandenen Karten werden mit der Karte der Grundstruktur „prokrustet“, wodurch sich die Koordinaten des neuen Objektes ergeben. Dieses für alle weiteren neuen Objekte identische Verfahren resultiert schlussendlich in einer kompletten Karte.

5.3. Resultate

Interpretierbarkeit der Wissenskarten. Die individuellen Ähnlichkeitsmatrizen wurden mit dem NMDS-Algorithmus RobuScal (Läge, Daub, Bosia, Jäger & Ryf, 2005) zweidimensional skaliert. Zur Überprüfung der statistischen Qualität der resultierenden NMDS-Karten wird der NMDS-Stress herangezogen (vgl. Borg & Groenen, 1997; Läge et al., 2005). Dieser Wert steht für die Abbildungsgüte der einer Karte zugrunde liegenden Ähnlichkeitseinschätzungen im Vergleich zu den relativen Verhältnissen der Objektpositionen in der Karte. Die Werte der individuellen NMDS-Lösungen (Grundstruktur und Vollstruktur) bewegen sich durchgängig unter den von Gigerenzer (1981, S. 353) als kritisch erachteten Niveaus bei 15 ($M = 0.099$, $SD = 0.033$) und bei 25 Objekten ($M = 0.122$, $SD = 0.031$). Die individuellen Wissenskarten bauen sich somit aus systematisch zueinander passenden Ähnlichkeitsurteilen auf. Damit ist eine wichtige Grundvoraussetzung gegeben, die relationalen Ähnlichkeitsurteile überhaupt als Messinstrument für das strukturelle Wissen einsetzen zu können.

Zeitliche Stabilität von Wissensstrukturen. Diese Auswertung dient dem Nachweis der ausreichenden Re-Test-Reliabilität von NMDS-Lösungen bei gleichem Wissensstand einer Person. Zu diesem Zweck werden die zwei NMDS-Lösungen aus den Erhebungszeitpunkten Ä1 und Ä2 aller Versuchspersonen miteinander verglichen. Neben den *intrapersonalen* Abweichungen zeigt ein solcher Vergleich auch die *interpersonalen* Unterschiede auf. Vergleicht man alle Karten durch multiple Prokrustes-Transformationen miteinander, ist es möglich, die relationalen Beziehungen (gemessen am AvgLoss) zwischen allen Personen und Zeitpunkten wiederum mit RobuScal zu skalieren und räumlich darzustellen (Prokrustesbasierte Personenskalierung, Läge, 2001). Abbildung 43 zeigt die Personenkarte als Ergebnis dieser Berechnungen. Pro Versuchsperson sind jeweils zwei Erhebungszeitpunkte in der Karte abgebildet (durch Striche miteinander verbunden). Die Normkarte ist als Kreuz eingetragen. Die grauen Punkte kennzeichnen die Karten zum Erhebungszeitpunkt Ä1, die schwarzen Punkte diejenigen zum Zeitpunkt Ä2. Die intrapersonalen AvgLoss-Werte sind ebenfalls in der Karte eingetragen.

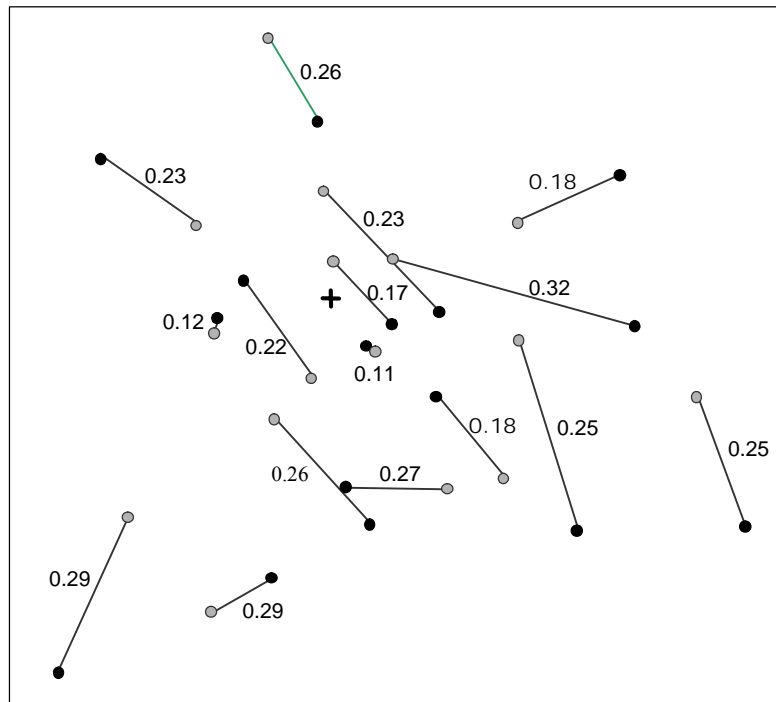


Abbildung 43. Multipler Kartenvergleich der beiden Erhebungszeitpunkte Ä1 und Ä2 mittels Prokrustesbasierter Personenskalierung: Mittelwert aller AvgLoss-Werte = 0.35, Min. = 0.12, Max. = 0.60.

Der mittlere AvgLoss der *intrapersonalen* Vergleiche beträgt $M = 0.23$ ($SD = 0.06$). Derjenige der *interpersonalen* Prokrustes-Transformationen, also aller möglichen Vergleiche zwischen den Versuchspersonen $M = 0.37$ ($SD = 0.08$). Der Stichprobenunterschied ist nicht durch den Zufall zu erklären ($t(494) = 1.648, p < 0.001$). Aufgrund der geringen Anzahl an intrapersonalen Messungen wurde zur Absicherung des inferenzstatistischen Maßes zusätzlich die Effektstärke berechnet. Auch diese bestätigt mit einem sehr hohen Wert von $d = 1.67$ das hoch signifikante Resultat (vgl. zur Einschätzung der Höhe von Effektstärken Bortz & Döring, 2002). Die Karten derselben Personen streuen merklich weniger als diejenigen zwischen den Probanden. Das gesetzte Ablehnungskriterium ist erfüllt.

Stabilität von Wissenskarten bei Erweiterung des Wissens. Ungeklärt ist, ob eine Integration von neuem Wissen ohne Akkomodationsleistungen der bestehenden Struktur (= Vorwissen) vollzogen werden kann. Das heißt, dass sich die Ordnung der Grundstruktur beim Einfügen von neuen Objekten stabil verhalten sollte, also nur marginale Positionsverschiebungen der Grundstrukturobjekte aufweist. Gewisse Unterschiede sind aufgrund der Verfahrenslogik zu erwarten (vgl. Einleitung).

Zur Bestimmung des strengeren Richtwerts der zu erwarteten Abweichungen wird folgendes theoretisches Maß herangezogen. Die Positionen der 15 Länder aus der mit allen Urteilen berechneten 25er Karte mit den Positionen der selben 15 Länder aus der sequentiell berechneten Karten werden miteinander über die Prokrustes-Transformation verglichen. In Abbildung 44 ist das Ausmaß der Veränderung anhand der berechneten Daten ersichtlich. Diese Abweichungen beschreiben methodisch betrachtet den Einfluss der neuen Objekte auf die Positionierung der Grundstrukturobjekte (hier: AvgLoss = 0.17). Die strukturelle Ordnung (Cluster-

zugehörigkeit) ist und die relationalen Beziehungen zwischen den einzelnen Objekten sind immer noch gewährleistet.

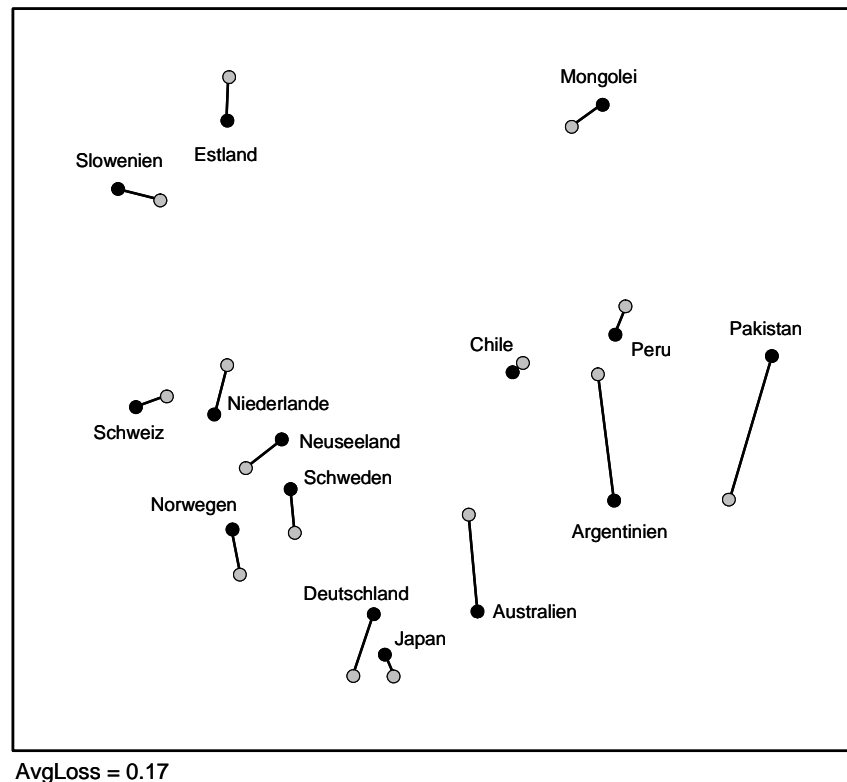


Abbildung 44. Abweichung der Grundstruktur: Prokrustes-Transformation der 15 Grundstruktureobjekte zwischen der sequentiell berechneten 25er Karte (grau) und der aus allen Urteilen berechneten 25er Karte (schwarz). AvgLoss = 0.17.

Im Gegensatz zu den berechneten Daten in Abbildung 44 können in den Matrizen der empirisch erhobenen Versuchspersonendaten Inkonsistenzen in den Ähnlichkeitsurteilen auftreten. Zu erwarten ist somit, dass die empirischen Daten gleich oder leicht über dem berechneten Wert von 0.17 zu liegen kommen. Die Versuchspersonendaten wurden mit dem identischen Verfahren – wie im vorigen Abschnitt beschrieben – miteinander verglichen. Mit einer durchschnittlichen Abweichung im AvgLoss von $M = 0.16$ ($SD = 0.07$) liegen die empirischen Werte im Mittel sogar etwas unter dem strengen Kriterium und weit unterhalb der etwas großzügiger angesetzten Grenze von 0.28.

Stabilität bei der Reduzierung von Ähnlichkeitsurteilen – Sequentielle Erweiterung. Die folgenden Auswertungen beschäftigen sich mit der Frage, ob über ein *sequentielles Einpassen* von Objekten hinlänglich mit der über die *Vollmatrix* erstellten NMDS-Lösungen vergleichbare Karten auf individueller Ebene produziert werden können (vgl. Tabelle 2). Der Unterschied im AvgLoss innerhalb derselben Personen beträgt im Mittel $M = 0.18$ ($SD = 0.07$). Unklar ist, ob dieses Niveau als Kosten für die dabei gewonnene Effizienz in der Diagnostik tolerierbar ist? Mit anderen Worten: Findet lediglich eine methodische Assimilation der relationalen Objektpositionen statt? Ein individuelles Beispiel soll der Veranschaulichung dienen. Abbildung 45 zeigt die strukturelle Abweichung in den Karten, basierend auf sequentieller Teilmatrix und auf der Vollmatrix der Versuchsperson 12. Diese Abweichung von 0.17 reprä-

sentiert den mittleren Differenzwert von 0.18. Die Karte weist keine groben Verletzungen der relationalen Beziehungen und der Clusterzugehörigkeit auf. Die beiden Strukturen sind praktisch identisch. Vergewenwärtigt man sich des Weiteren zum einen als rechnerische Vergleichsbasis die Abweichung von $\text{AvgLoss} = 0.3$, wenn die Merkmalswerte um $\pm 10\%$ vom Originalwert (als hypothetisches Maß für temporäre Wissensschwankungen und Gewich- tungsunterschiede) und die daraus generierten Ähnlichkeitswerte um ± 1 (als hypothetisches Maß für das Rauschen im Urteilsverhalten) verändert werden, erhalten wir in den empirischen Daten einen um fast 50% geringeren Wert. Auch das in der Einleitung festgesetzte Kriterium im AvgLoss von 0.28 wird weit unterschritten.

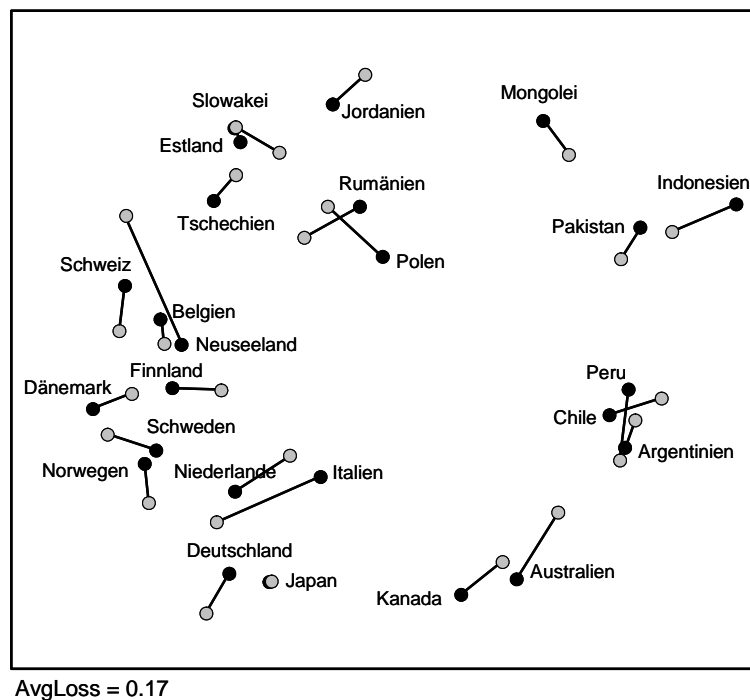


Abbildung 45. Prokrustes-Transformation der Karten von Versuchsperson 12, berechnet auf der Basis der Vollmatrix (schwarz) und der sequentiellen Einpassung (grau). AvgLoss = 0.17.

Als erstes zentrales Kriterium dient auch hier die Personenspezifität. Wiederum sollen die Abweichungen aus den Kartenvergleichen zwischen der Bedingung Vollmatrix und sequen- tieller Teilmatrix innerhalb derselben Person merklich kleiner ausfallen als zwischen ver- schiedenen Versuchspersonen. Zur Erinnerung: Der mittlere AvgLoss der *intrapersonalen* Vergleiche beträgt $M = 0.18$ ($SD = 0.07$). Derjenige der *interpersonalen* Prokrustes- Transformationen, also aller möglichen Vergleiche zwischen den Versuchspersonen, beträgt $M = 0.39$ ($SD = 0.08$). Im t -Test wird der Unterschied auf dem 0.1%-Niveau signifikant ($t(494) = 1.647, p < 0.001$). Die sehr hohe Effektstärke von $d = 2.51$ unterstreicht dieses deut- liche Ergebnis.

Der zweite zu erfüllende Grenzwert zur Einschätzung der Güte der vorhandenen Abweichung ergibt sich über den Vergleich mit der aufgrund von Messwiederholungen zu erwartenden Schwankung. Die Abweichung einer Person zwischen den zwei Erhebungszeitpunkten Ä1 und Ä2 dient als Richtwert. Abbildung 46 zeigt das Ergebnis auf individueller Ebene.

Zu sehen ist, dass die intrapersonalen Kartenabweichungen zwischen zwei Erhebungszeitpunkten bis auf drei Ausnahmen (Versuchspersonen 1, 4 und 14) größer sind als die resultierenden Abweichungen aus den zwei Skalierungsverfahren. Der mittlere AvgLoss aus der Messwiederholung beträgt $M = 0.23$ ($SD = 0.06$), derjenige aus der Datenreduktion $M = 0.18$ ($SD = 0.07$). Dieser Unterschied ist mit einer hohen Effektstärke von $d = 0.79$ auf dem 1%-Niveau signifikant ($t(15) = 1.753, p < 0.01$).

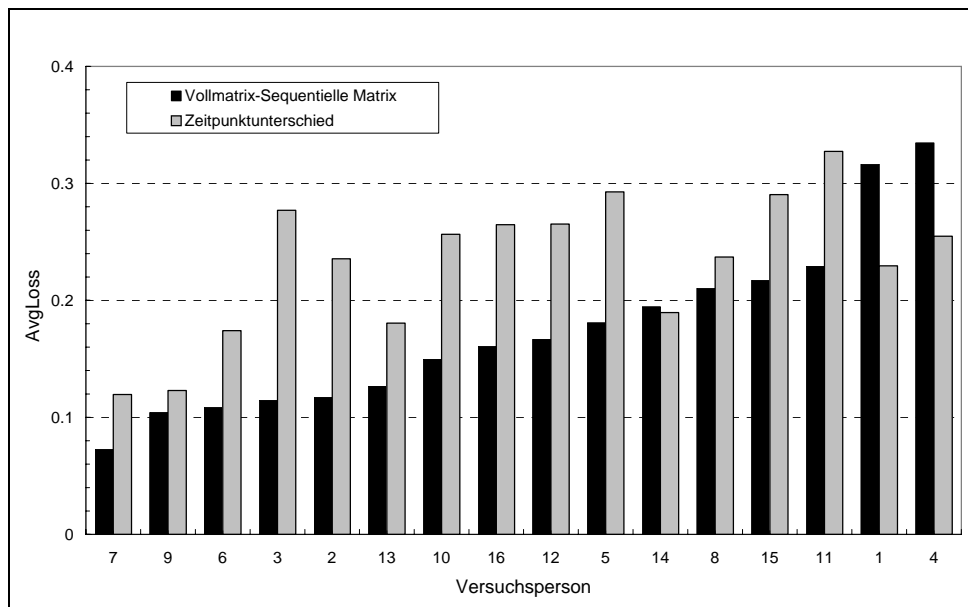


Abbildung 46. Individueller Vergleich zwischen Datenreduktion und Schwankungen von Messwiederholungen. Die Daten auf der Abszisse sind in eine Rangfolge gebracht worden, beginnend mit der kleinsten Abweichung im AvgLoss zwischen Vollmatrix und sequentieller Teilmatrix (schwarze Balken). Die individuellen Messwiederholungsdifferenzen ($\bar{A}1$ - $\bar{A}2$) sind mit grauen Balken aufgeführt.

Stabilität bei der Reduzierung von Ähnlichkeitsurteilen – Reduktionsstufen. Durch das Verfahren der *Reduktionsstufen* wird zusätzlich untersucht, wie viele Urteile zwischen den neuen Objekten und den 15 Objekten der Grundstruktur notwendig sind, um die relationalen Äquivalenzen der Objekte zu bewahren. Eine Computersimulation diente der Datengenerierung, sowohl für berechnete als auch für die empirischen Ähnlichkeitsurteile (Abbildung 47). In 12 Reduktionsstufen⁹ wurden die Kartenabweichungen für eine steigende Anzahl von Lücken berechnet (Reduktionsstufe 3 bedeutet beispielsweise, dass pro neues Objekt zufällig drei Ähnlichkeitswerte fehlen). Als Kontrollbedingung diente die aus der Vollmatrix berechnete NMDS-Lösung. Die Abweichungen der Grundstrukturobjekte und der neuen Objekte sind getrennt aufgeführt.

Betrachten wir zuerst die berechneten Daten: Hierbei handelt es sich um mathematisch exakte Proximitäten und dementsprechend sind keine Inkonsistenzen in den Urteilen enthalten. Die

⁹ Auf die Reduktionsstufen 13, 14, 15 wurde verzichtet, da in einem zweidimensionalen euklidischen Raum drei Bezugspunkte notwendig sind, um einen neuen Punkt in eine Objektkonfiguration eindeutig einzupassen.

gefundenen Differenzen in der Kartenstruktur können folglich der Skalierungs-Methodik und deren Umgang mit lückenhaften Daten zugeschrieben werden. Aufgrund des sequentiellen Einfügens der neuen Objekte, werden die Positionen der Grundstrukturobjekte kaum mehr verändert (s. Abbildung 47: „berechnet 15 GS-Objekte“) und bleiben auf dem Niveau von 0.18. Je mehr Lücken in den Zeilen der neuen Objekte gesetzt werden, desto größer wird die Abweichung im Vergleich zur Kontrollbedingung (s. Abbildung 47: „berechnet 10 neue Objekte“) und erreicht nach Ausschöpfung der mathematisch möglichen Reduzierung den relativ hohen Wert von über 0.4.

Die empirischen Daten der Differenzen der neuen Objekte aufgrund der Reduktionsstufen weichen in Abbildung 47 im Vergleich zu den berechneten Daten erwartungsgemäß etwas mehr von der Kontrollbedingung ab. Die Erwartungshaltung begründet sich dadurch, dass zusätzlich zu den methodisch bedingten Abweichungen hier noch Inkonsistenzen im individuellen Urteilsverhalten zum Tragen kommen. Werden Lücken gesetzt, beeinflussen solche Inkonsistenzen die Struktur (= das Skalierungsverfahren) in höherem Maß. Die sequentiell berechnete Karte der empirischen Daten kommt bei der Reduktionsstufe 0 (= keine zusätzlichen Lücken gesetzt) im Mittel – wie bereits vorher erwähnt – unter die Messwiederholungsdifferenzen zu liegen (s. Abbildung 47: „Zeitpunktunterschiede“). Die neuen Objekte überschreiten jedoch schon ab der dritten Reduktionsstufe das gesetzte Gütekriterium der Zeitpunktunterschiede zwischen \bar{A}_1 und \bar{A}_2 von $\text{AvgLoss} = 0.23$ (s. Abbildung 47: „empirisch 10 neue Objekte“). Die neuen Objekte überschreiten jedoch schon ab der dritten Reduktionsstufe das gesetzte Gütekriterium der Zeitpunktunterschiede zwischen \bar{A}_1 und \bar{A}_2 von $\text{AvgLoss} = 0.23$ (s. Abbildung 47: „empirisch 10 neue Objekte“). Die Objekte der Grundstruktur stagnieren demgegenüber auf demselben Niveau von 0.18 (s. Abbildung 47: „empirisch 15 GS-Objekte“). Hier zeigen sich keine beziehungsweise vernachlässigbare Unterschiede zwischen simulierten und empirischen Daten.

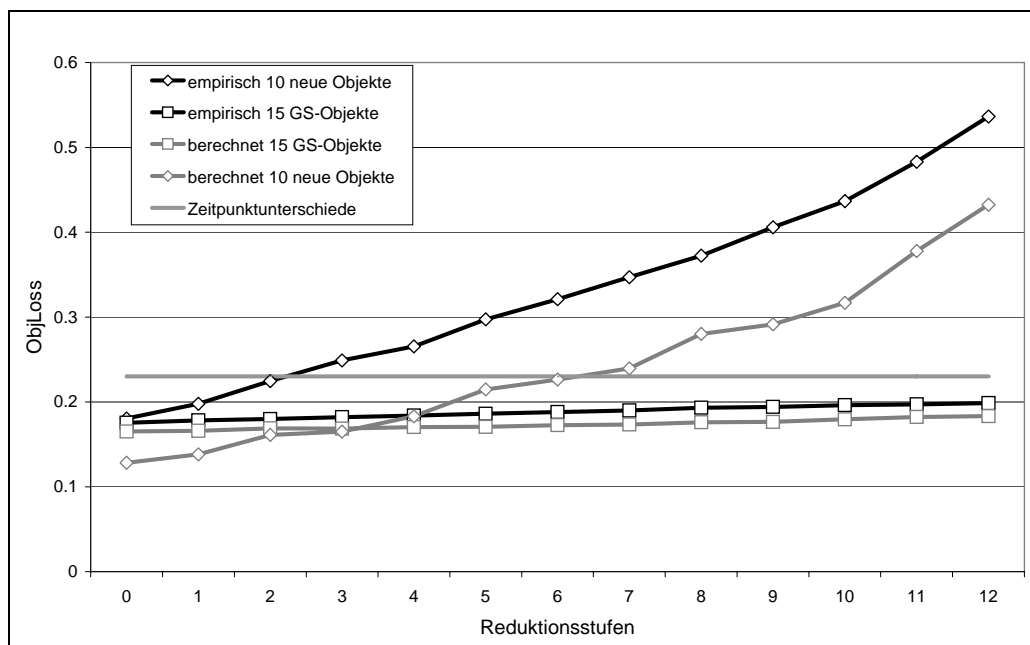


Abbildung 47. Gemittelte Simulationsdaten der Reduktionsstufen mit empirischen und berechneten Daten. Nähere Erläuterungen im Text.

5.4. Diskussion

Zeitliche Stabilität. Die Ergebnisse in Abbildung 43 haben gezeigt, dass die gelernten Strukturen über die Zeit zwar nicht identisch, aber sicherlich personenspezifisch bleiben. Die Veränderungen zwischen den einzelnen Karten pro Versuchsperson sind geringer als die Unterschiede zwischen den Versuchspersonen. Dies, obwohl die verschiedenen Versuchspersonen auf dieselbe Struktur hin mit demselben Lernmaterial gelernt haben, die einzelnen Strukturen eng um die berechnete Normstruktur variieren und somit eine hohe Homogenität aufweisen. Umso erfreulicher, dass trotz dieses Umstandes die Idiosynkratie der Lösungen erhalten bleibt.

Trotzdem sind die mittleren intraindividuellen Abweichungen von $\text{AvgLoss} = 0.23$ nicht gleich Null, die Karten somit nicht in dem Ausmaß deckungsgleich, wie man es vielleicht erwarten hätte können. Ein möglicher Erklärungsgrund liegt im Experimentalsetting. Wie in Abbildung 42 ersichtlich, brachte das Versuchsdesign mit sich, dass die erste Erhebung (Ä1) im Anschluss an eine Lernphase durchgeführt wurde (zeitlicher Abstand ca. 5 Minuten). Die zweite Erhebung (Ä2) hingegen isoliert zu einem späteren Zeitpunkt stattfand. Dadurch könnte zum ersten Erhebungszeitpunkt Wissen aus dem „Kurzzeitgedächtnis“ in die Urteile eingeflossen sein. Ein Vergessenseffekt kann somit nicht ausgeschlossen werden, der für eine noch stabilere Grundstruktur sprechen würde, wir also die wahre Reliabilität des Verfahrens unterschätzen. Vor diesem Hintergrund betrachtet kann die beschriebene Personenspezifität durchaus als Indiz für eine über die Zeit stabile Repräsentation von kognitiven Strukturen gewertet und die Reliabilität des Verfahrens als Instrument für die Sachwissensdiagnostik als gegeben betrachtet werden.

Stabilität bei Erweiterung des Wissens. Die Positionen der 15 Objekte der Grundstruktur verändern sich im Raum, sobald weitere Objekte in die Berechnung der Karte mit einfließen. Das Ausmaß der Veränderung bei der theoretischen Berechnung ist mit einem AvgLoss von 0.17 als relativ gering zu bezeichnen. In der abgebildeten Karte (Abbildung 44) ist ersichtlich, dass die Positionen und strukturellen Relationen zwischen den verschiedenen Objekten weiterhin gewährleistet bleiben. Das Einfügen von weiteren Objekten in eine bestehende Struktur hat unter den gegebenen Umständen keine wesentliche akkomodative Veränderung der Grundstruktur zur Folge. Ausgehend von stabilem Vorwissen sind somit lediglich als assimilative Prozesse zu bezeichnende Feineinstellungen der Methodik zu beobachten. Die Positionsverschiebungen bei den empirischen Daten sind im Mittel mit 0.16 sogar geringer als diejenigen der berechneten Daten. Demzufolge scheint auch bei empirischen Daten keine merkliche Umstrukturierung stattzufinden.

Eine plausible Erklärung für die besseren Ergebnisse der empirischen Daten ist im Versuchsdesign zu vermuten. Die Versuchspersonen mussten zuerst in mehreren Lernsituationen die Objekte der Grundstruktur lernen. Die Latenzphasen zwischen den Lerneinheiten erlauben es der Person, (unbewusst) differenzierte Ordnung in die Wissensinhalte zu bringen, was in letzter Konsequenz zu einem konsistenten Urteilsverhalten und stabiler Objektpositionierung führt. Das einmalige Lernen neuen Wissens verursacht kurzfristig zwar eine bessere Kartenqualität. Abstriche sind dann aber im Urteilsverhalten zu suchen (vgl. dazu Kapitel 2 in dieser

Arbeit). Der NMDS-Algorithmus RobuScal berücksichtigt bei der Objektpositionierung zueinander passende Ähnlichkeitsurteile stärker und Schwankungen in der Konsistenz schwächer. Somit hätten die neuen Objekte weniger Einfluss auf die Positionen der Grundstrukturobjekte. Die aus den realen Daten berechneten – perfekt konsistenten – Urteile weisen jedoch keine solchen Gewichtungsunterschiede auf, weder innerhalb der Grundstruktur noch in der erweiterten Objektmenge. Der Einfluss der neuen Objekte auf die Positionierung ist äquivalent zu jenen der Grundstruktur. Die Ergebnisse zeigen, dass neue Objekte eine gewisse Veränderung der Objektpositionen mit sich bringen, dass das Ausmaß dieser Veränderung strukturell jedoch zu vernachlässigen ist.

Des Weiteren folgt, dass mit einer Erhöhung der Objektzahl in der Basisstruktur diese Stabilität aufgrund des RobuScal-Algorithmus ansteigen wird. Eine Eigenschaft der NMDS ist es, dass Kognitive Karten bei steigender Objektzahl und gleich bleibender Dimensionalität durch die Einschränkung der Freiheitsgrade bezüglich der Objektpositionierung an struktureller Stabilität gewinnen. Monte-Carlo Studien haben gezeigt, dass die untere Grenze für strukturell stabile Wissenskarten bei 15 Objekten liegt (Läge et al., 2005 und im Zusammenhang mit dem Zweck der Erkennung von Wissensdefiziten Kapitel 4 in dieser Arbeit). Dementsprechend ist bei der Verwendung einer größeren Objektmenge für die Grundstruktur sogar eher eine Steigerung der Stabilität zu erwarten. Die Richtigkeit dieser Hypothese lässt sich quantitativ untermauern, wenn die Objektmenge der Grundstruktur erhöht wird (s. Tabelle 3). Bereits 5 Objekte mehr in der Grundstruktur lassen den AvgLoss zwischen der sequentiell berechneten und der auf der Vollmatrix beruhenden Karte auf ein sehr tiefes Niveau von 0.11 fallen.

Tabelle 3. Steigerung der Stabilität durch Erhöhung der Objektzahl der Grundstruktur (simulierte Daten). Nähere Erläuterungen im Text.

	AvgLoss
15 GS-Objekte + 10 Neue Objekte	0.17
20 GS-Objekte + 10 Neue Objekte	0.11
25 GS-Objekte + 10 Neue Objekte	0.11

Zu beachten ist, dass es sich bei der Grundstruktur um eine repräsentative Objektauswahl des interessierenden Wissensbereiches handeln muss. Die später einzufügenden Objekte dürfen keine grundlegenden Neuerungen (d.h. zum einen keine neuen Merkmalsdimensionen und zum anderen keine Vergrößerung des Wertebereichs der Merkmale) in die Wissenslandschaft einbringen, um akkomodative Umstrukturierungen zu vermeiden.

Sequentielle Einpassung und Reduktionsstufen. Das Verfahren der sequentiellen Einpassung scheint geeignet, um neue Objekte in eine bestehende Wissensstruktur zu integrieren, ohne dass über den Verzicht auf die Ähnlichkeitsurteile innerhalb neu gelernten Objekte merkliche strukturelle Veränderungen in den Kognitiven Karten zu beobachten sind. Fraglich ist, ob die absolute Abweichung von der Kontrollbedingung akzeptabel ist oder eine zu starke Veränderung der Struktur stattfindet?

Als erstes Gütemaß wurde die Personenspezifität herangezogen. Dabei wird ersichtlich, dass die Unterschiede von zwei Karten mit unterschiedlichen Urteilmengen der gleichen Person

geringer sind als jene zwischen zwei beliebigen Karten. Diese Personenspezifität spricht für die Möglichkeit, mit weniger Urteilen eine strukturell vergleichbare Karte zu erstellen. Eine aussagekräftige Beurteilung des notwendigen Toleranzbereiches kann jedoch noch nicht durchgeführt werden. Diese Schwierigkeit, die Abweichung von der Kontrollkarte auf ihre strukturelle Tolerierbarkeit zu beurteilen, wurde daraufhin mit dem Vergleich zu normalen intrapersonalen Schwankungen angegangen. Als normale Schwankungen gelten in diesem Fall die strukturellen Abweichungen, die sich bei einer Messwiederholung ergeben. Diese Zeitpunktunterschiede müssen bei dieser Methodik als Toleranzbereich akzeptiert werden. Demzufolge würden Abweichungen, die sich in diesem Bereich befänden noch keine verfälschende strukturelle Veränderung ausmachen, da sie auch durch den Messzeitpunkt entstehen können (zweites Gütemaß). Die intrapersonalen Abweichungen zwischen der Kontrollbedingung Vollmatrix und sequentiell errechneter Karte sind signifikant geringer als die interpersonalen Zeitpunktunterschiede.

Für die praktische Anwendung der ähnlichkeitsbasierten Sachwissensdiagnostik bedeutet dies, dass in Zukunft die Qualitätsbeurteilung der Erweiterung von Basiswissen lediglich über den Vergleich der neuen Objekte mit denjenigen der Ankerstruktur angegangen werden kann und komplett auf die Erhebung paarweiser Ähnlichkeiten innerhalb neuen Wissens verzichtet werden kann. Damit ergibt sich ein Effizienzgewinn in der Datenerhebung, welcher sich bei steigender Anzahl an neu gelernten Objekten mehr und mehr auszahlt.

Aufgrund dieser Ergebnisse wäre es nicht abwegig anzunehmen, dass nicht alle Vergleiche zwischen einem neuen Objekt und den Objekten der Grundstruktur notwendig sind, um eine tolerierbare Positionierung zu ermöglichen. In der zusätzlich durchgeführten Simulation der verschiedenen Reduktionsstufen hat sich gezeigt, dass bei relationalen Daten, im Gegensatz zu metrischen, weitaus mehr Vergleiche notwendig sind, um den Zerfall der Struktur zu verhindern. Somit ist es nicht möglich, sich dem geometrischen Minimum von drei Urteilen für jedes zusätzliche Objekt anzunähern. Im Gegenteil, es konnten in dieser Studie lediglich drei der 15 Vergleiche (= 20%) weggelassen werden, um im gewählten Toleranzbereich von Messwiederholungen zu bleiben. Außerdem war das verwendete Kriterium der Zeitpunktunterschiede eher zu hoch angesetzt, womit Effizienzsteigerung 2 in dieser Studie nicht empfohlen werden kann.

Dieses Resultat erscheint jedoch in einem positiveren Licht, wenn man eine wesentliche Eigenschaft der NMDS-Methodik in Betracht zieht. NMDS-Karten gewinnen bei steigender Objektzahl und gleich bleibender Skalierungsdimensionalität an struktureller Stabilität aufgrund der damit verbundenen Einschränkung der Positionierungsfreiheitsgrade. Bei der vorliegenden Untersuchung wurde mit 15 Objekten das Minimum für strukturell stabile Karten verwendet (Läge, Daub, Bosia, Jäger & Ryf, 2005). Bei der Verwendung von mehr Grundstrukturen steigt zwar der Erhebungsaufwand für dieses Basiswissen aufgrund der exponentiell von der Objektzahl abhängigen Urteilsmenge, im weiteren Verlauf könnte dafür der Effizienzgewinn bei der Beurteilung neuer Wissensinhalte diesen Mehraufwand kompensieren. Und, die Objektpositionierung des erweiterten Wissens sollte aufgrund der erhöhten Stabilität der Ankerstruktur an Validität gewinnen.

Einordnung der Befunde in den Gesamtzusammenhang. Diese Studie zeigt, dass die Diagnose von merkmalsbasiertem Objektwissen über das Verfahren der *Sequentiellen Erweiterung* effizienter durchgeführt werden kann als mit den herkömmlichen paarweisen Ähnlichkeitsurteilen zwischen allen Objekten (*Vollerhebung*). Das entwickelte Verfahren der Sequentiellen Erweiterung basiert auf dem lerntheoretisch begründeten Assimilationsprozess (vgl. Ausubel, 2000). Voraussetzung für dieses Verfahren ist eine stabile Grundwissensstruktur, in die das neue Wissen integriert werden kann, ohne den vorhandenen Bezugsrahmen zu verändern. Neben dem zugrunde liegenden, stabilen Merkmalswissen bezüglich der einzelnen Objekte der Grundstruktur sind die zeitliche Stabilität von kognitiven Strukturen und ein Lernen und Integrieren von neuem Wissen ohne methodische und kognitive Akkomodationsleistung der bestehenden Struktur zwingend notwendig und anhand der vorliegenden Resultate auch gegeben.

Das ursprüngliche, extrem ineffiziente Diagnostikkonzept von vollständigen, paarweisen Ähnlichkeitsurteilen passt sich in der Folge an diese methodischen Rahmenbedingungen an und fordert einen gestuften, modularen Lernprozess. Der Lerner muss zu Beginn eine globale und solide Ankerstruktur bilden. Vorzugsweise kontrolliert man diesen Prozess darüber, dass stichprobenartig ein repräsentativer Auszug an Objekten und charakterisierenden Merkmalsdimensionen aus dem gesamten Wissensumfang vermittelt wird. Zukünftig hinzukommendes Faktenwissen lässt sich dann in assimilativer Weise integrieren, ohne dass die Objektbeziehungen der Grundstruktur wesentlichen Umstrukturierungskräften unterliegen. Nach und nach lässt sich der Wissensumfang bis zu einem expertennahen Niveau über einen für die lernende Person optimal konstruierten Lernprozess erweitern und vertiefen.

Dieses Vorgehen wird auch in Modellen der Lernpsychologie, welche einen strukturierten und kontrollierten Lernprozess im Sinne einer assimilativen Verknüpfung neuer Information mit bereits Gespeichertem propagieren (z.B. Novak, 2002, beziehend auf Ausubel, 2000; Piaget, 1976). Der Grund liegt in der angestrebten Vermeidung von akkomodativen Anpassungsprozessen beim Lernen. Jedes Neuaufnehmen von Informationen erfordert eine Neustrukturierung des Gesamtbildes im Sinne einer Integration und Elaboration der erhöhten Wissensmenge. Dies erfordert einen kognitiven Aufwand, der größer oder kleiner sein kann, je nachdem ob das neue Wissen mit dem bereits Gespeicherten in Einklang steht oder nicht. Ein einfach konstruiertes, prototypisches Beispiel soll dies verdeutlichen. Angenommen, eine Person weiß über verschiedene Länder wie groß sie sind, wie viele Menschen in den Ländern leben und was das mittlere individuelle Einkommen der jeweiligen Nation ist. Dieser Merkmalssatz konstituiert einen Satz an stabilen relationalen Beziehungen. Nun lernt diese Person von einem ihr bislang unbekannten Land dieselben Merkmalssätze. Ohne größeren Aufwand gelingt es, diese neue Information in das Gedächtnis einzubauen, denn es kommen lediglich neue Ausprägungsgrade auf den Merkmalsdimensionen hinzu und die relationale Verknüpfung zum bestehenden Wissen kann als reines „Ergänzen“ betrachtet werden. Dies ist ein klassischer Assimilationseffekt im Sinne Piagets (1976). Dieselbe Person erfährt nun von anderer Seite, dass es in Ländern unterschiedliche Regenmengen oder Durchschnittstemperaturen gibt. Diese Erweiterung der Dimensionalität des kognitiven Raumes verursacht Reibungen. Die Auflösung dieser Dissonanzen erfordert eine Umstrukturierung und Neubeurteilung

der relationalen Beziehungen zwischen den Gedächtnisinhalten und führt zu einer sich von der alten wesentlich unterscheidenden Wissensstruktur. Dieser Prozess wird als Akkomodation bezeichnet (Piaget, 1976). Es liegt auf der Hand, dass es für den Lerner einfacher ist, wenn neues Wissen assimilativ integriert werden kann oder in Ausubel's (2000) Worten, dass man den Lerner dort abholt, wo er sich gerade befindet.

Die motivationale und didaktische Ausrichtung, Gestaltung und Adaptierbarkeit von (virtuellen) Lernumgebungen auf die Bedürfnisse des Lerners fördern den Lernerfolg unbestritten erheblich. Durch ein effizientes Diagnoseverfahren kann genügend Zeit für die eigentliche Wissensvermittlung aufgewendet werden. Neben der Adaptierbarkeit der Wissensvermittlung an die Bedürfnisse des Lerners kann dem Lerner durch die Adaptivität einer intelligenten Diagnostik der richtige Lerninhalt präsentiert werden (vgl. Leutner, 1992a, 1995). Die strukturelle Wissensdiagnose anhand Kognitiver Karten ermöglicht es zusätzlich zur Adaptierbarkeit des Systems, adaptiv auf das Wissen des Lerners zu reagieren und ihm maßgeschneidert die noch nicht gewussten Lerninhalte anzubieten. Außerdem misst das Verfahren der NMDS nicht unzusammenhängende Merkmalssätze, sondern fokussiert explizit auf relationale Beziehungen und genügt somit dem Umstand, dass sich Sachwissen durch Vernetztheit, Komplexität und Differenziertheit auszeichnet (s. Mandl & Fischer, 2000).

Anwendung kann dieses Verfahren vor allem im computerbasierten Lernen (eLearning) beziehungsweise in Intelligenten Tutoriellen Systemen (ITS) finden. Dabei liegt der Fokus weniger auf ebenfalls wichtigen Aspekten wie der Motivation und / oder der didaktischen Aufbereitung und Vermittlung, sondern vielmehr auf der Diagnostik von Wissen, wobei der durch das Tutorielle System vermittelte Wissenszuwachs im Mittelpunkt steht. Dabei ist zu beachten, dass die benötigte Zeit zur Diagnose der Lernzeit abgeht. Aufgrund dieses Dilemmas zwischen dem zeitlichen Aufwand der Wissensvermittlung und jenem der Wissensdiagnose (vgl. Leutner, 1992b) konzentrieren sich die meisten Bemühungen auf die ebenfalls wichtige instruktionalistische Seite. Die Resultate dieser Studie eröffnen nun einen neuen Zugang zur effizienteren Diagnose von Sachwissen mittels der NMDS-Methodik, wodurch ein Beitrag zur Verringerung der erwähnten Problematik geliefert werden kann.

5.5. Literatur

- Ausubel, D. P. (2000). *The acquisition and retention of knowledge*. Dodrecht: Kluwer Academic Publishers.
- Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling – Theory and applications*. New York: Springer.
- Bortz, N. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Commandeur, J. J. F., & Heiser, W. J. (1993). *Mathematical derivations in the proximity scaling (PROXSCAL) of symmetric data matrices (Tech. Rep. No. RR-93-03)*. Leiden, The Netherlands: Department of Data Theory, Leiden University.
- Der Fischer Weltalmanach (2002). *Staaten. Der Fischer Weltalmanach – Online Datenbank*. Aufgerufen 3. Juli 2002 unter http://www.weltalmanach.de/staat/staat_liste.html
- Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie*. München: Reinhardt.
- Gower, J. C. & Dijksterhuis, G. B. (2004). *Procrustes Problems*. New York: Oxford University Press.

- Läge, D. (2001). *Ähnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Läge, D., Daub, S. Bosia, L., Jäger, C. & Ryf, S. (2005). *Die Behandlung ausreißer-behafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag*. AKZ-Forschungsbericht Nr. 21. Zürich: Angewandte Kognitionspsychologie.
- Lee, K., Byatt, G. & Rhodes, G. (2000). Caricature effects, distinctiveness, and identification: testing the face-space framework. *Psychological Science*, 11, 379-385.
- Leutner, D. (1992a). *Adaptive Lehrsysteme. Instruktionspsychologische Grundlagen und experimentelle Analysen*. Weinheim. Psychologie Verlags Union.
- Leutner, D. (1992b). Das Testlängendilemma in der lernprozess-begleitenden Wissensdiagnostik. *Zeitschrift für Pädagogische Psychologie*, 24, 232-248.
- Leutner, D. (1993). Das gleitende Testfenster als Lösung des Testlängendilemmas: Eine Robustheitsstudie. *Zeitschrift für Pädagogische Psychologie*, 7, 33-45.
- Leutner, D. (1995). Adaptivität und Adaptierbarkeit multimedialer Lehr- und Informationssysteme. In L. J. Ising & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia* (S. 140-149). Weinheim: Psychologie Verlags Union.
- Mandl, H. & Fischer, F. (Hrsg.). (2000). *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*. Göttingen: Hogrefe.
- Marx, W., & Hejj, A. (1989). *Subjektive Strukturen*. Göttingen: Hogrefe.
- Marx, W., & Läge, D. (1995). *Der ideologische Ring*. Göttingen: Hogrefe.
- Novak, J. D. (2002). Meaningful learning: the essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, 86 (4), 548-571.
- Piaget, J. (1976). *Die Äquilibration der kognitiven Strukturen*. Stuttgart: Klett.
- Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least square method with optimal scaling features. *Psychometrika*, 42, 7-67.
- Van Merriënboer, J. J. G., Van Den Berg, K. G. & Maaswinkel, D. M. (1989). Some experiences with two intelligent tutoring systems for teaching computer programming: PROUST and LISP-tutor. In J. M. Pieters (Ed.), *Intelligent tutorial systems and instruction* (pp. 7-44). Enschede: OTG Onderwijsleerprocessen.

6.

**Adaptive Lernempfehlungen
über Prokrustes-
Transformation und
Kohonennetz**

Zusammenfassung

Vielschichtige Anforderungen an die individuelle Ausbildung erfordern eine flexible Wissensakkumulation. Computerbasierte Lehrsysteme erlauben es, adaptiv auf den Wissensstand des Einzelnen Rücksicht zu nehmen, insofern dass über nicht-linear strukturierte Wissensvermittlung möglichst effizient spezifische Defizite erkannt und beseitigt werden. Grundlage für die inhaltliche Adaptivität eines solchen Intelligenten Tutoriellen Systems (INTUS) ist eine umfassende, relationale Diagnostik des individuellen, deklarativen Faktenwissens. Über die Nonmetrische Multidimensionale Skalierung und Prokrustes-Transformation werden mangelhaft gewusste Elemente detektiert. Kohonen-Netze ermöglichen dann die Abschätzung der einzelnen Merkmalsdefizite und führen zu einer individuell angepassten Zusammenstellung der Lerninhalte. Eine Simulation (Daub, 2001) zeigte eindrucklich, dass dieses didaktische Konzept effizienter zu Expertise führt als ein linear strukturierter Lernprozess. Ein empirisches Vergleichsexperiment konnte infolge mangelhafter Variablenkontrolle die Resultate nicht replizieren. Der hohe Aufwand zur Gewährleistung umfassender Voraussetzungen für die nochmalige empirische Überprüfung lässt eine Erweiterung der Simulation als fruchtbarer erscheinen. Es wird ein simulatives Konzept vorgestellt, welches Erkenntnisse aus der Gedächtnisforschung implementiert mit dem Zweck, sich realitätsnah verhaltende, künstliche Lerner kompetitiv zu vergleichen.

Schlüsselwörter: Sachwissen, Merkmalswissen, Wissenskarte, Kognitive Karte, Kognitive Struktur, Nonmetrische Multidimensionale Skalierung, Prokrustes-Transformation, Kohonen-Netz, Lernexperiment, Empirische Simulation, Adaptive Lernempfehlungen, Intelligentes Tutorielles System

6.1. Einleitung

Adaptive Lernempfehlungen. Die Anforderungen an die berufliche Kompetenz des Einzelnen haben sich in den letzten Jahrzehnten drastisch verändert. Reichte noch vor nicht so langer Zeit der Abschluss einer Ausbildung mehrheitlich aus, um sich für den Rest des Lebens im Arbeitsmarkt zu halten, muss in der heutigen Gesellschaftsstruktur ein hohes Maß an Flexibilität und Anpassbarkeit an den Tag gelegt werden. Die daraus folgende notwendige Wissensakkumulation hat sich im Alltag mit dem Schlagwort „lebenslanges Lernen“ durchgesetzt. Damit steigen auch die Leistungen, die Bildungseinrichtungen (wie Fachhochschulen oder Universitäten) bieten müssen. Es liegt jedoch in der Sache selbst, dass Institutionen, welche große Massen durch die Ausbildung schleusen müssen, eine gewisse Starrheit im Hinblick auf die Befriedigung individueller Lernerbedürfnisse aufweisen. Ein Dozent / Lehrer wird aus Gründen des ökonomischen Lehrbetriebes nur auf das Gesamt des streuenden Lernerwissens Rücksicht nehmen können, womit für gute Schüler Zeit ungenutzt bleibt und schlechte überfordert sind.

Mit dem Aufkommen und der Weiterentwicklung von Computern bietet sich demgegenüber ein beachtliches Potential für den flexiblen Einsatz solcher Systeme als Lehr- und Lernmedi-

um. Ein optimal konstruiertes Lernprogramm kann – um nur einige unvollständige Vorteile zu nennen – auf individuelle Defizite eingehen, ist von Ort und Zeit unabhängig, nimmt Rücksicht auf individuelles Lerntempo, kann jederzeit unterbrochen und aufgenommen werden usw. Mit den immer besser werdenden Leistungen des Computers ergeben sich auch differenziertere Möglichkeiten zur Entwicklung eines qualitativ hochwertigen, tutoriellen Systems. Angefangen von ersten Lehrmaschinen anfangs des 20. Jahrhunderts (Pressey, 1926), gefolgt von Weiterentwicklungen durch Skinner (1958) und Crowder (1959), mündete, mit dem Aufkommen der Künstlichen Intelligenz (K.I.), der Fortschritt in Ansätzen, welche sich als Intelligente oder Adaptive Tutorielle Systeme (ITS, ATS) bezeichnen lassen. Ein wichtiger Vorteil solcher Systeme ist auf der einen Seite die Adaptierbarkeit und auf der anderen Seite die Adaptivität. Adaptierbarkeit meint, dass sich das Medium an den Lerner anpassen bzw. der Lerner selbst Einfluss auf das System nehmen kann, sei dies z.B. bezogen auf die Abfolge der Lerneinheiten, Modifikationsmöglichkeiten auf der Benutzeroberfläche usw. Adaptivität dagegen meint, dass das System auf der Basis einer validen Wissensdiagnostik den Lerner dort abholt, wo er sich gerade befindet. Das tutorielle System bietet dem Lerner denjenigen Schwierigkeitsgrad an, welcher bewältigt werden kann und stellt diejenige Information zusammen, die noch mangelhaft gewusst wird (zur Definition von adaptiven Lehrsystemen und einem ausführlichen Überblick über historische Entwicklungen von automatisierten Lehr- und Lernsystemen s. Leutner, 1992a.).

Solch ein Verfahren ist, ergänzend zu den vorher angesprochenen Vorteilen, maximal effizient, da Mehrfachlernen vermieden werden kann und nicht einem linear strukturierten Lehrplan unterliegt und beschleunigt den Lernprozess unbestritten erheblich. Dem kritischen Leser wird aber aufgefallen sein, dass man es mit der postulierten Effizienzsteigerung mit einem zweischneidigen Schwert zu tun hat: Auf der einen Seite muss, um maximale Adaptivität (bezogen auf die Auswahl der Lerninhalte) gewährleisten zu können, eine umfassende und exakte Statusdiagnose des Lernerwissens vollzogen werden. Zum anderen darf diese Messung nicht allzu viel Zeit in Anspruch nehmen, da die Zeit zur Diagnose logischerweise der potentiellen Lernzeit abgeht (Dieses Problem wurde von Leutner, 1992b als „Testlängendilemma“ bezeichnet). Vor allem, wenn deklaratives Sachwissen vermittelt werden will, müsste für eine gesamtheitliche, aber für optimale Adaptivität eben notwendige Diagnostik jedes einzelne Faktum in unzähligen Einzelfragen abgetestet werden. Eine stichprobenartige Ziehung, wie dies in den meisten Prüfungen umgesetzt ist, verringert zwar den zeitlichen Aufwand, läuft aber Gefahr, das Wissen des Lerners nur „oberflächlich“ zu messen und dementsprechend eine unvollständige Auswahl an Lernmaterial zu liefern. Im Folgenden wird eine Verfahrenslogik vorgestellt, die es ermöglichen soll, den zeitlichen Diagnoseaufwand in einem vertretbaren Rahmen zu halten, aber dennoch Wissen umfassend und exakt zu diagnostizieren.

Wissensdiagnostik. Das deklarative Merkmalswissen einer Person über eine Anzahl von Objekten lässt sich mit so genannten „Kognitiven Karten“ darstellen (Läge, 2001, Marx & Hejj, 1989; Marx & Läge, 1995). Als dimensionale Gebilde zeigen sie die Relationen, die eine Person zwischen diesen Objekten sieht, in Form von kleineren oder größeren Distanzen auf und lassen als Gesamtbild damit Rückschlüsse auf die von der Person als wesentlich erachteten Merkmalsbereiche zu (s. Abbildung 48). Relationale Beurteilungen einer Person über einen

Satz von Objekten bilden die Datengrundlage und lassen sich mit der Methode der Nonmetrischen Multidimensionalen Skalierung (NMDS) in ein Raummodell überführen (Borg & Groenen, 1997). Diese globalen Ähnlichkeitseinschätzungen zwischen Objekten werden auf der Basis des „dahinterliegenden“ Merkmalswissens gebildet (vgl. Klauer, 1989) und sind verhältnismäßig expertisenunabhängig, wobei von einem Experten das Urteil natürlich qualitativ besser ausfallen sollte. Eine Ähnlichkeitsbeurteilung erhebt also das gesamte Wissen über ein Element und ist dadurch effizient.

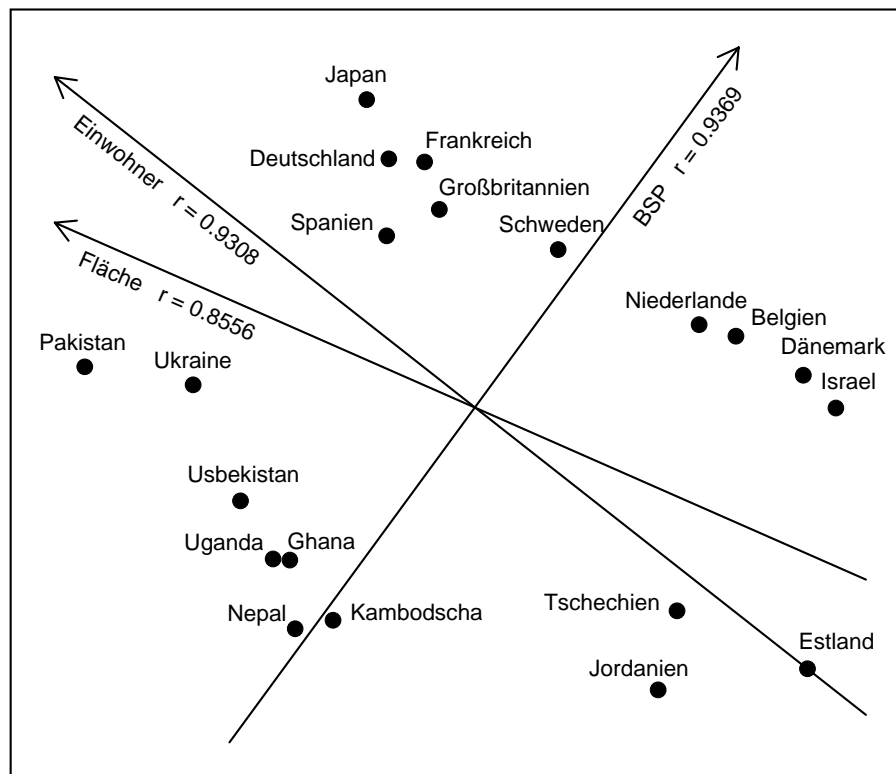


Abbildung 48. NMDS-Karte, welche im Experiment als Expertenmodell Verwendung findet (vgl. Punkt Methode). Die zugrunde liegenden Merkmalsbereiche sind Bruttonationaleinkommen, Einwohnerzahl und Fläche, welche über multiple Regression mit genügend hohen Korrelationen in die Karte eingepasst werden konnten.

Gleichzeitig erlauben Kognitive Karten, das Wissen über ein spezifisches Objekt durch deren Position in der Gesamtstruktur nachzuvollziehen und die Wissensqualität zu bewerten (Prokrustes-Transformation, s. Abbildung 49; vgl. Gower & Dijksterhuis, 2004): Wir konnten zeigen, dass sich Lernfortschritte auf der Gesamtstrukturebene nachzeichnen lassen (vgl. Kapitel 2 in dieser Arbeit). Es existieren zudem starke Hinweise darauf, dass unter bestimmten Voraussetzungen Vorhersagen über die Wissensqualität auf der Einzelobjektebene gemacht werden können (vgl. Kapitel 4 in dieser Arbeit).

Bis zum jetzigen Zeitpunkt kann das System mangelhaft gewusste Objekte detektieren. In einem weiteren Schritt soll nun darauf abgezielt werden, das Merkmalswissen, welches „hin-

ter“ den Ähnlichkeitsurteilen steht, zu schätzen¹⁰. Als mögliches Verfahren bieten sich Kohonen-Netze an (nach ihrem Erfinder Teuvo Kohonen benannt, vgl. Kohonen, 1982, 1984, 1995), welche eine Interpretation Kognitiver Karten auf Merkmalsebene erlauben. Kohonen-Netze sind selbstorganisierende Merkmalslandschaften und gehören zur Familie der Neuronalen Netze. Kohonen-Netze werden mit den korrekten Merkmalsausprägungen trainiert. Jedes Objekt eines Objektbereiches findet im trainierten Netz ein Neuron, welches seinen Merkmalsvektoren am besten entspricht. Auch so entsteht also eine Struktur, welche die Ähnlichkeitsrelationen der Objekte abbildet. Die Lernerstruktur kann nun mittels Prokrustes-Transformation auf das trainierte Netz gelegt werden.

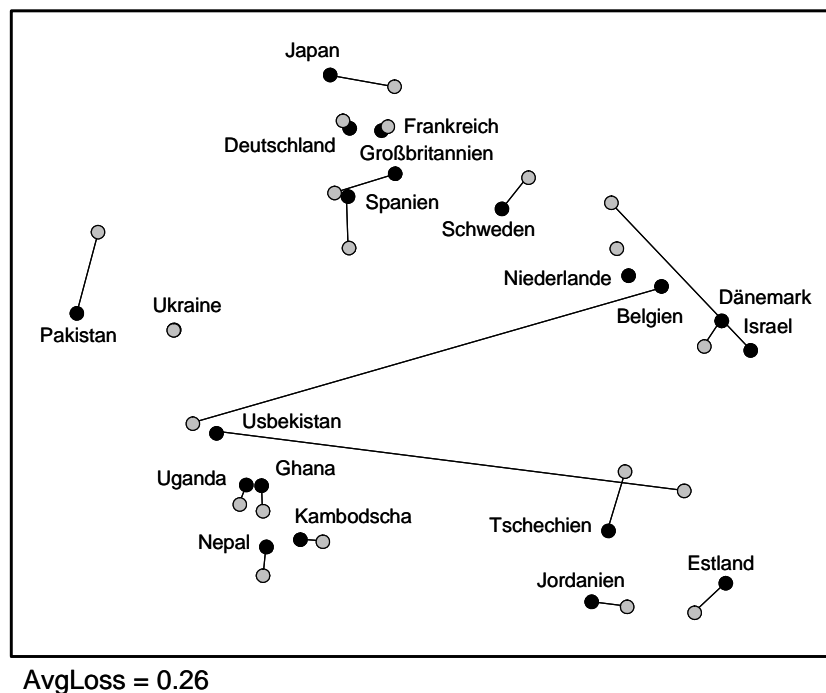


Abbildung 49. Ergebnis einer Prokrustes-Transformation. Schwarz dargestellt sind die Positionen in der Expertenkarte, grau diejenigen eines (fiktiven) Lerners. Ohne größeren mathematischen Aufwand springt ins Auge, welche Objekte der Lerner gut kennt (d.h. korrekt verortet) und welche nicht (hier: Usbekistan, Belgien, Israel). Die Gesamtabweichung, welche sich aus dem Mittel aller Einzelobjektabweichungen (ObjectLoss oder kurz: ObjLoss), beträgt in diesem prototypischen Beispiel AverageLoss (AvgLoss) = 0.26.

Im Gegensatz zu NMDS-Karten, innerhalb derer lediglich Abweichungen festgestellt werden können, haben Kohonen-Netze den Vorteil, dass alle Neuronen (= jede Position im Raum) eine Merkmalsvektorkombination aufweisen, womit bei einem von der korrekten Position abweichenden Objekt die Merkmalsvektoren abgelesen werden können. Auf der Basis dieser Abweichungen erhält man eine Schätzung des Lerner-Merkmalswissens, so dass die größten

¹⁰ Die nun folgenden Ausführungen lehnen sich an die Arbeiten von Daub (2001) an. Der interessierte Leser, welcher sich eingehender mit den einzelnen methodischen Details und der theoretischen Einbettung des Verfahrens auseinandersetzen will, sei darum auf diese Arbeit verwiesen.

Differenzen ausgewählt und für die kommende Lektion zusammengestellt werden können¹¹. Dieses zweistufige Verfahren ist in Abbildung 50 schematisch dargestellt.

Fraglich bleibt, wie hoch die Fehlerquelle bei der Überführung einer Kognitiven Karte in ein Kohonen-Netz ist. Zu diesem Zweck hat Daub (2001) eine Simulation durchgeführt, in der überprüft wurde, ob das Endresultat eines trainierten Netzes bezogen auf die Positionen der Objekte in genügend hohem Masse mit der Position in der NMDS-Karte (externes Kriterium) übereinstimmt. Als Inhaltsgebiet fungierte geographische Information über Länder (Bruttosozialprodukt, Einwohnerzahl, Fläche). Das Ziel war es, die Parameter-Einstellungen des Kohonen-Netzes so zu wählen, dass am Ende des Trainings Netz und Kognitive Karte bestmöglich übereinander passen. Die Extraktion und Skalierung der Objektpositionen im Kohonen-Netz, verglichen mit den Positionen der Kognitiven Karte (skaliert auf der Basis von berechneten Ähnlichkeitsmassen) über die Prokrustes-Transformation ergab den kleinen Wert von $\text{AvgLoss} = 0.16$, womit von einer hohen Übereinstimmung der beiden Repräsentationsformen ausgegangen werden kann (vgl. Daub, 2001, S. 87).

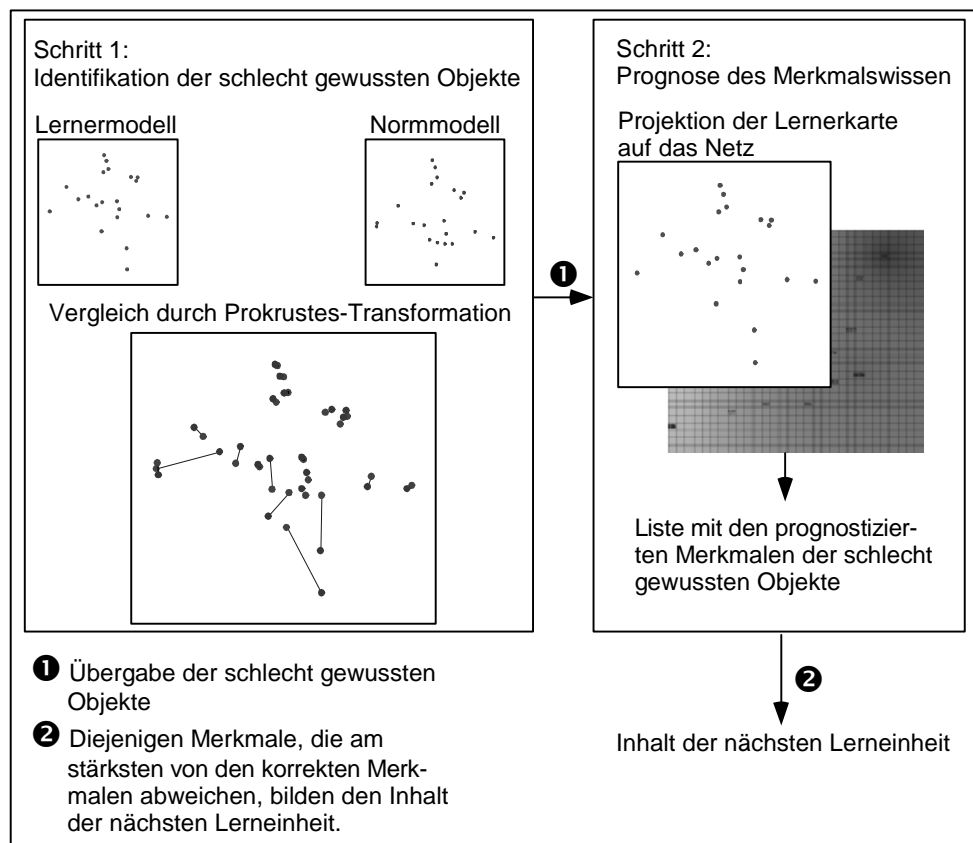


Abbildung 50. Zweistufiges Diagnostikverfahren über Prokrustes-Transformation und Kohonen-Netz (aus Daub, 2001, S. 171).

¹¹ Zu beachten: Es wird nicht behauptet, dass aufgrund der Abweichungen in der Position eines Objektes im Neuronalen Netz der Lerner genau über dieses vom Kohonen-Netz geschätzte Merkmalswissen verfügt. Dies wäre gleichbedeutend mit Falschwissen. Ebenso kann eine Abweichung dadurch entstehen, dass ein Lerner über kein Wissen verfügt. Für den Einsatz als Diagnostikverfahren wird diese Unterscheidung jedoch hinfällig, da sowohl Falschwissen als auch lückenhaftes Wissen als Wissensdefizit interpretiert wird.

Effizienz bei der adaptiven Auswahl von Lehreinheiten in einer kompetitiven Simulation. Wir haben gesehen, dass die beiden Repräsentationsformen Kognitive Karte und Kohonen-Netz ineinander überführbar sind und somit die Schätzung des Merkmalswissens einer Person auf der Basis von relationalen Urteilen möglich ist. Diese beiden Schritte können nun, wie in Abbildung 50 gesehen, kombiniert und auf ihre Qualität hin an *Künstlichen Lernern* überprüft werden. Daub (2001) entwickelte eine Simulation, in der getestet wurde, wie schnell ein virtueller Laie zu einem Experten wird (Die Verfahrenslogik wird mit „INTUS“ bezeichnet).

Dazu wurden mehrere Lernzyklen simuliert. In jedem Zyklus wird aus Ähnlichkeitsurteilen der momentane Wissensstand des Lerners auf der Basis einer NMDS-Karte geschätzt. Das Expertenmodul diagnostiziert daraufhin die Wissensdefizite des Lerners. Am Ende der Diagnostik steht dann eine Liste mangelhaft gewusster Objekte (aufgrund der Prokrustes-Transformation: ObjLoss) und eine Schätzung des Merkmalswissens des Künstlichen Lerners über genau diese Objekte (aufgrund des Kohonen-Netzes: Merkmalsvektoren). Für die Lerneinheit sollen zum einen diejenigen Merkmale ausgewählt werden, deren Abweichung zu den tatsächlichen Merkmalsausprägungen maximal ist und zum anderen, dass genau so viele Merkmale ausgewählt werden, wie für die nächste Lektion vom Umfang her möglich sind. Dieser Zyklus wird solange durchlaufen, bis entweder die Anzahl an diagnostizierten Objekten oder die Kartenqualität des Lerners einen bestimmten Schwellenwert unterschreiten.

Durchgeführt wurde die Simulation mit 21 Ländern, definiert durch die Merkmale Bruttonationaleinkommen, Einwohnerzahl und Fläche. Lernerwissen wurde manipuliert, indem in jedem Simulationsdurchgang (insgesamt 25 Durchgänge = 25 Künstliche Lerner) bei 17 der 21 zwei der drei Merkmale in zufälliger Richtung verändert wurden (im Maximum um Faktor 20). Damit lag ein mittleres Geographiewissen vor, eine ideale Voraussetzung für die tutorielle Intervention. Ungeachtet didaktischer oder lerntheoretischer Annahmen wurde festgelegt, dass nach jeder Lerneinheit die darin enthaltenen 5 Merkmalsinformationen zu seinem bisherigen Wissen perfekt beherrscht wurden (Lernrate = 100%; Vergessensrate = 0%).

Um die Leistung des Verfahrens einer kompetitiven Evaluation zugänglich zu machen, wird ein Vergleichsmodell benötigt. Gewählt wurde ein „Lehrbuchlernen“. Diese vereinfachte Sichtweise kennzeichnet sich durch eine Aufteilung des gesamten Wissensstoffes in Lektionen, wobei jedes Element genau einmal und in festgelegter Reihenfolge vermittelt wird.

Als Qualitätsmasse zur Evaluation der INTUS-Performanz verwendete Daub (2001, S. 175) drei Kriterien:

- *Inhaltliche Qualität der Lerneinheit:* Anzahl Merkmale in jeder Lerneinheit, welche auch tatsächlich das falsche Wissen verbessern (= Trefferrate).
- *Aktueller Wissensstand des Lerners:* Momentaner Gesamtfehler im Wissen der Künstlichen Lerner. Es wird sowohl berücksichtigt, wie viele Merkmale der Lerner noch falsch in seinem mentalen Lexikon gespeichert hat, als auch, wie stark diese von den korrekten Werten abweichen.
- *Kognitive Karte des Lerners:* Abweichung der aktuellen Kognitiven Karte des Lerners vom Normmodell (AvgLoss).

Berichtet werden im Folgenden lediglich die Resultate zum ersten und dritten Kriterium. Die inhaltliche Qualität der Lerneinheit lässt sich anhand der Trefferrate beschreiben. Je mehr schlecht gewusste Merkmale diagnostiziert werden, desto besser die Auswahl.

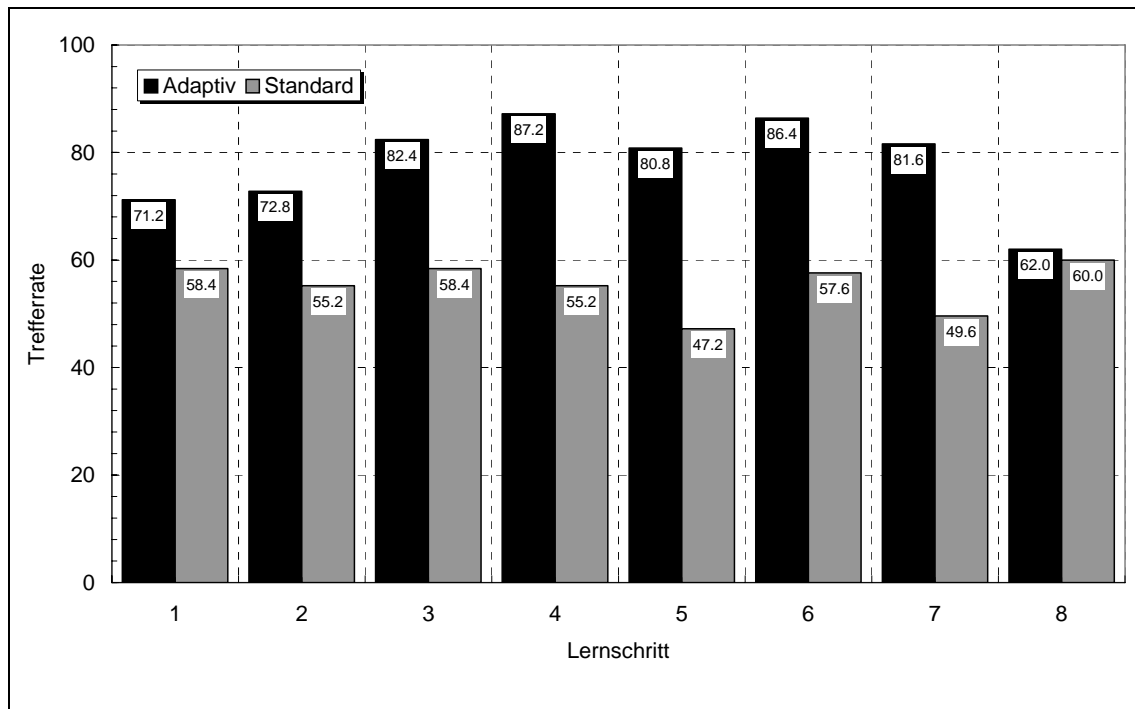


Abbildung 51. Verhältnis von neuer Information zur gesamten Information einer Lerneinheit (nach Daub, 2001, S. 177).

Abbildung 51 ist zu entnehmen, dass über alle Lernschritte der INTUS-Lerner (*Adaptiv*) mehr Informationen erhält, die sein Wissen verbessern als der Lehrbuch-Lerner (*Standard*) und beträgt im Durchschnitt ca. 80%. Zwischen Lernschritt drei und sieben ist der Unterschied am höchsten, was dadurch zu erklären ist, dass die Adaptivitätslogik dann besonders gut greift, wenn ein bestimmtes, mittleres Vorwissen vorhanden ist. Der Unterschied wird dann aber immer kleiner, da INTUS nun bereits die größten Fehler korrigiert hat (dies entspricht dem zweiten, hier nicht berichteten Kriterium), womit der Lehrbuch-Lerner den anfänglichen Rückstand aufholen kann.

Die Verbesserung der Karten ist in Abbildung 52 zu sehen. Der Adaptiv-Lerner gleicht seine Wissensstruktur viel schneller und kontinuierlicher an als der Standardlerner und erreicht nach sieben Lernschritten ein praktisch vernachlässigbares Niveau. Es ist anzumerken, dass die Kartenabweichung nicht das zentrale Evaluationskriterium darstellt. Viel wichtiger ist, ob INTUS die großen Wissensdefizite detektiert, was, wie Daub (2001) eindrücklich zeigen konnte, gegeben ist.

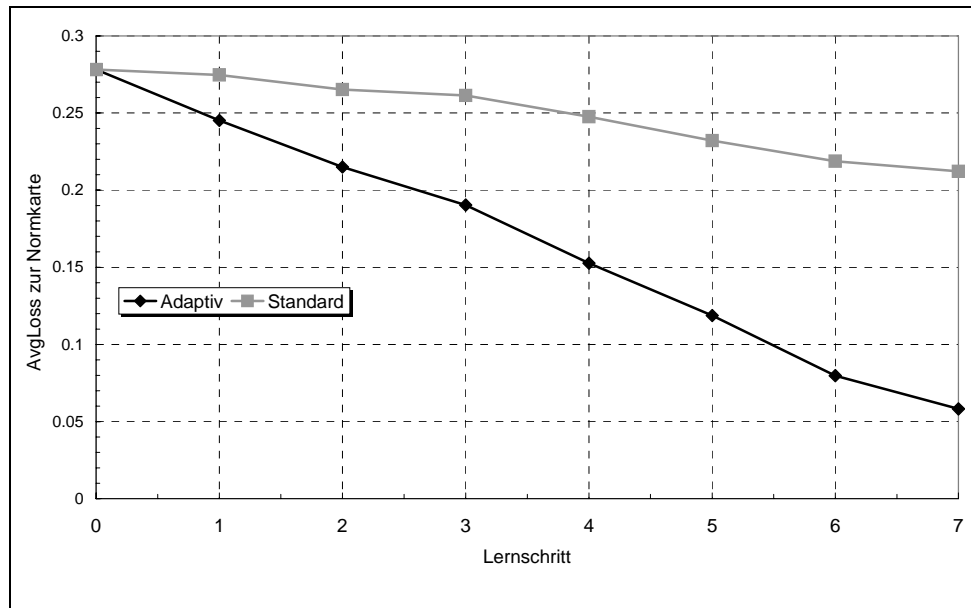


Abbildung 52. AvgLoss der Lernerkarten zur Normkarte (nach Daub, 2001, S. 181).

Damit ist zumindest die prinzipielle Tauglichkeit des Verfahrens aufgezeigt worden. Diese Vorgehensweise leidet jedoch an der stark konstruierten Ausrichtung. Zu kritisieren ist, dass hier ein realer Lernvorgang nur bedingt simuliert wurde, da, wie vorher erwähnt, immer perfektes Lernen ohne Vergessen stattfand. Wie Daub (2001) ebenfalls bemerkt, muss die Verfahrenslogik anhand von Lerner aus Fleisch und Blut überprüft werden. Dabei wird bei den Probanden innerhalb einer Lerneinheit wohl nicht alles gelernt werden und zudem auch einige Information wieder aus dem Gedächtnis verschwinden, wenn einige Zeit zwischen den Lern- und Diagnosesitzungen verstreicht. Dies sollte aber, so die Erwartungen, den Adaptivitätsvorteil sogar noch vergrößern, da ein Buchlerner, welcher die Information nur einmal zu Gesicht bekommt, keine Möglichkeit hat, einmal Vergessenes erneut zu korrigieren.

Zu diesem Zweck wurde ein Lernexperiment durchgeführt, welches vom Prinzip her exakt gleich konstruiert war wie die Simulation. Der INTUS-Lerner (*Adaptiv*) erhält die Zusammenstellung der Lektion aufgrund der Diagnostik, der Schulbuchlerner (*Standard*) bearbeitet eine festgelegte Reihenfolge. Einzige Ausnahmen waren die Anzahl Merkmale pro Lerneinheit (zwölf im Gegensatz zu fünf in der Simulation), die Objektmenge (20 anstatt 21) und das Vorwissensniveau (nicht nur Falschwissen wie in der Simulation, sondern auch lückenhaftes Wissen).

Erwartete Befunde. Analog der Resultate aus der Simulation sollte das INTUS-Verfahren die mangelhaft gewussten Merkmale beim INTUS-Lerner besser detektieren (operationalisiert über höhere Trefferrate verglichen mit der Zufallsauswahl beim Standard-Lerner) und zu einer schnelleren Angleichung der Kognitiven Karten an das Normmodell führen (operationalisiert über AvgLoss). Der aktuelle Wissensstand des Lernalters kann nicht analog zur Simulation bestimmt werden, da dafür eine kontinuierliche, externe Wissensdiagnostik (z.B. anhand Multiple-Choice) notwendig wäre. Um die Versuchspersonen nicht zusätzlich mit dem zeitraubenden Ausfüllen von Fragebögen zu belasten, musste darauf verzichtet werden.

6.2. Methode

Stichprobe. 35 Versuchspersonen im Alter zwischen 21 und 54 Jahren ($M = 26.3$, $SD = 6.1$) nahmen am Lernexperiment teil. Die Probanden (18 w, 17 m) verfügen entweder über eine Hochschulausbildung, eine Berufslehre, die Matura oder sind Universitätsstudenten.

Lernmaterial. Das zu lernende Wissensgebiet umfasst 20 Länder mit den Merkmalsdimensionen Bruttosozialprodukt pro Kopf in US\$, Einwohnerzahl und Fläche in km^2 (20 Länder mal 3 Merkmalsdimensionen = 60 Merkmale). Die Daten stammen aus dem Fischer Weltatlas (2002). Die Auswahl folgte dem Kriterium der Repräsentativität, das heißt, dass solche Länder ausgewählt wurden, welche auf den Merkmalsdimensionen möglichst unterschiedliche Ausprägungen aufwiesen und somit dem kognitiven Raum gut „aufspannen“. Um Vorwissen auszuschließen, wurden den realen Nationenbezeichnungen fiktive Namen zugeordnet (s. auch Punkt *Coverstory*): Troxa (Belgien), Jellis (Dänemark), Brikka (Deutschland), Ayuri (Estland), Verrosa (Frankreich), Budan (Ghana), Sarnin (Großbritannien), Ordal (Israel), Zelles (Japan), Lissipia (Jordanien), Erul (Kambodscha), Lorsis (Nepal), Solete (Niederlande), Madesch (Pakistan), Dalurien (Schweden), Falmir (Spanien), Xenex (Tschechien), Kolluri (Uganda), Reskebra (Ukraine) und Gremino (Usbekistan). Um das Vorwissen zu generieren, wurden sieben Objekte komplett richtig gelernt, zwei Objekte komplett falsch und zwei Objekte wurden komplett nicht gelernt. Vier weitere Falschinformationen und vier weitere Lücken waren auf je vier Objekte verteilt. Die Merkmalsausprägungen der zwei komplett falsch gelernten Objekte sind diejenigen der zwei komplett nicht gelernten. Die restlichen vier Falschinformationen bewegen sich innerhalb des minimalen und maximalen Merkmalsbereichs, wobei ein korrekter kleiner Wert sehr groß und ein großer sehr klein gemacht wurde.

Coverstory. Zur Beantwortung der Fragestellung ist zwingend notwendig, dass eine hohe Kontrolle über das Faktenwissen gewährleistet ist. Um Vorwissen auszuschließen, wurden anstelle der realen Nationen fiktive Namensbezeichnungen verwendet (vgl. vorher). Es war für die experimentelle Zielstellung wichtig, dass die Probanden über falsches Wissen verfügen, womit die Gefahr einer Konfundierung des Lernmaterials mit allfälligem Vorwissen ausgeschlossen werden musste. Eine Voruntersuchung ergab keinen Einfluss der neu konstruierten Namen auf assoziativ bedingte, implizite Tendenzen in den Merkmalsausprägungen. Um ethische Bedenken bezogen auf das anfängliche Lernen von falscher Information auszuschalten, wurde das Experimentalsetting mittels einer Coverstory anschaulicher gemacht: Die Probanden schlüpfen in die Rolle einer internationalen Kontrollinstanz eines fiktiven Kontinents, welche die Aufgabe bekam, nach Beendigung eines langjährigen kontinentalen Krieges Informationen über die verschiedenen Staaten zu sammeln. Es wurde ihnen mitgeteilt, dass es wichtig ist, die Informationen miteinander in Beziehung zu setzen, da für kommende Allianzgespräche potentiell zueinander passende Nationen sich anhand der Ähnlichkeiten ergeben. Das Wissen wird von Informanten aus den jeweiligen Regionen übermittelt. Im ersten Teil des Experimentes war die Information noch unvollständig, was auf die chaotischen Zustände in den Krisengebieten zurückzuführen sei. Der zweite Teil des Experimentes bestand dann darin, diese Lücken und das – wie sich nachträglich herausstellen sollte – aus dem Munde von sabotierenden Spionen stammende Falschwissen auszumerzen. Die Abgabe der Ähnlichkeits-

urteile stand in dieser Geschichte für die Zusammenstellung der gesammelten Informationen als Empfehlungsgrundlage für nachfolgende Verhandlungsgespräche zwischen den einzelnen Staaten.

Experimentalsetting: In Abbildung 53 ist schematisch der Experimentalablauf dargestellt. Die erste Sitzung diente dem Lernen des Vorwissens, welches sowohl richtige und falsche Informationen beinhaltete und unvollständig war (im Gegensatz zur Simulation musste Wissen zuerst generiert und nicht bestehendes Wissen verfälscht werden; s. auch Punkt Lernmaterial). Um mit der Ähnlichkeitsbeurteilung vertraut zu werden, beurteilten die Versuchspersonen zuerst 8 Früchte, anschließend erledigten sie die erste Lernphase (45 von 60 Merkmalen), gefolgt von einer Ähnlichkeitsbeurteilung aller 20 Objekte (zum Zwecke der Strukturbildung). Die erste Sitzung wurde von einer erneuten Lernphase und einem Merkmals-Fragebogen (Multiple-Choice) abgerundet. Zu Beginn der zweiten Sitzung repetierten die Versuchspersonen das Vorwissen und gaben Ähnlichkeitsurteile ab. Auf der Basis des Fragebogens und der ersten Ähnlichkeitsurteile der zweiten Sitzung wurden zwei homogene Experimentalgruppen gebildet: Zum einen die *Adaptiv-Gruppe* ($n = 17$), welche das Lernmaterial (12 Merkmale) auf der Basis des Diagnostikverfahrens erhielt (vgl. Abbildung 54), zum anderen die *Standard-Gruppe* ($n = 18$), welche pro Lerneinheit eine Zufallsauswahl (12 Merkmale) aus den 60 Merkmalssätzen vorgelegt bekam. Die Standard-Gruppe bekam somit über das ganze Experiment jedes Merkmal einmal zu Gesicht (5 Lerneinheiten mal 12 Merkmale = 60 Merkmale).

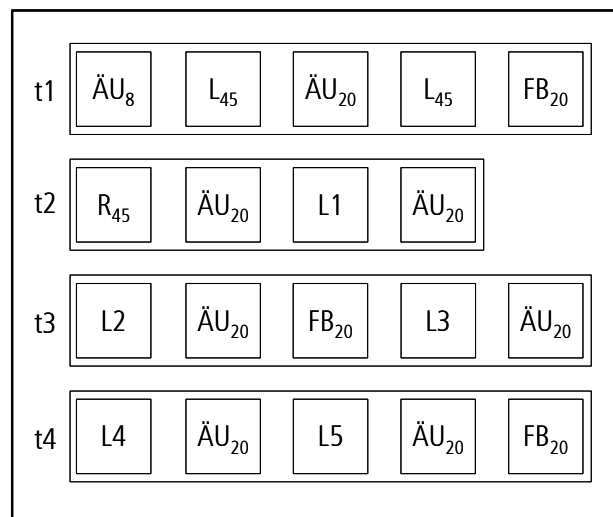


Abbildung 53. Experimentalsetting. t1 bis t5 bezeichnen die fünf Sitzungen im Abstand von je einer Woche. ÄU = Ähnlichkeitsurteile; L = Lerneinheit (insgesamt 5 Lektionen); FB = Fragebogen; R = Repetition. Die Indizes stehen für die Anzahl Objekte, welche getestet beziehungsweise für die Anzahl Merkmale, welche gelernt wurden.

Der Rest des Experimentes charakterisiert sich über den Wechsel zwischen Lerneinheit und Diagnoseeinheit, wobei für die Adaptiv-Gruppe jeweils die vorangegangenen Ähnlichkeitsbeurteilung als Grundlage für die Zusammenstellung der folgenden Lerneinheit darstellte. Ganz zum Schluss bearbeiteten die Probanden nochmals einen Multiple-Choice-Fragebogen zur komplementären Kontroll-Überprüfung des Merkmalswissens.

Diagnoseapparat. In Abbildung 54 ist der Diagnoseapparat schematisiert (s. auch Abbildung 50). Die Abgabe der Ähnlichkeitsurteile ermöglichte die Berechnung einer NMDS-Karte. Über die Prokrustes-Transformation ergaben sich für die Adaptiv-Gruppe eine Vorauswahl an mangelhaft gewussten Ländern (sechs höchste ObjLoss-Werte). Die Überführung der NMDS-Karte auf die Kohonen-Repräsentation erlaubte dann auf der Basis der somit gewonnen Merkmalsvektoren innerhalb der Vorauswahlobjekte eine Abschätzung der Differenzen zwischen korrekter Kohonenausprägung und Lernerausprägung auf Merkmalsebene. Die zwölf größten Differenzen führten zur Zusammenstellung der nachfolgenden Lerneinheit. Die Standardgruppe erledigte jeweils die Ähnlichkeitsdiagnose zur Kontrolle, die Auswahl des Lernmaterials geschah aber über eine Zufallsauswahl und war für alle Standard-Lerner gleich.

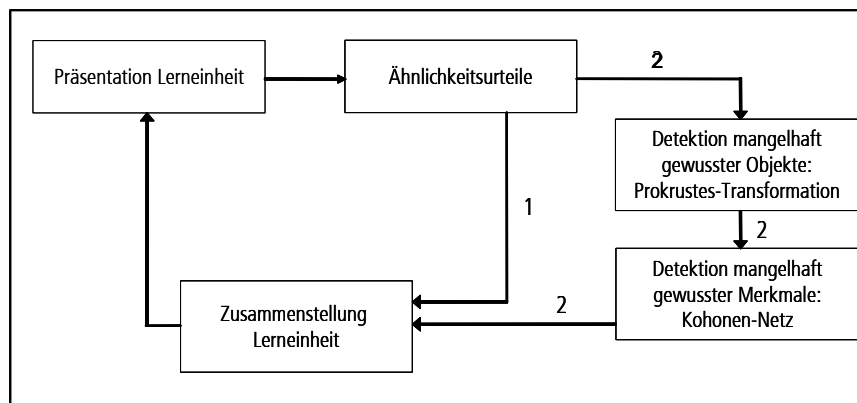


Abbildung 54. Diagnoseapparat. Die Adaptiv-Gruppe (2) erhielt die Lerneinheit auf der Basis von NMDS-Karte, Prokrustes-Transformation und Kohonen-Netz, die Standard-Gruppe (1) eine Zufallsauswahl. Nähere Erläuterungen im Text.

Normmodell. Als Berechnungsgrundlage zur Erstellung des Normmodells dienten die realen Merkmalsausprägungen pro Land. Über die untenstehende Formel ergab sich für jedes Länderpaar ein Unähnlichkeitsquotient, welcher mit dem NMDS-Algorithmus RobuScal (Läge, Daub, Bosia, Jäger & Ryf, 2005) zweidimensional skaliert werden konnte. Zwei Normmodelle wurden berechnet. Zum einen ein Normmodell auf der Basis der vorhandenen 45 Merkmalsätze des Vorwissens, zum anderen das endgültige Normmodell basierend auf der korrekten Merkmalsinformation (s. Abbildung 48). Das erste Normmodell diente zur Kontrolle der Qualität des gelernten Vorwissens (1. Sitzung und 1. Ähnlichkeitsbeurteilung der 2. Sitzung). Das definitive Normmodell fand dann Verwendung für die adaptive Diagnostik und als Lernkontrolle auf Strukturebene über alle restlichen Lernsitungen.

$$U\ddot{A}_{x,y} = \frac{|(x_i - y_i)/(x_i + y_i)| + |(x_j - y_j)/(x_j + y_j)| + |(x_k - y_k)/(x_k + y_k)|}{3}$$

Wobei: $U\ddot{A}_{x,y}$: Unähnlichkeitsquotient zwischen zwei Objekten
 x, y : Länderindex
 i : Bruttosozialprodukt pro Kopf in US\$
 j : Einwohnerzahl
 k : Fläche in km²

6.3. Resultate

Inhaltliche Qualität der Lerneinheit. Eine optimale Lerneinheit vermittelt nur diejenigen Merkmale, über die der Lerner fehlerhaftes Wissen verfügt. Die Qualität wird also über die Anzahl Merkmale, die anhand von Prokrustes-Transformation und Kohonen-Netz ausgewählt werden und die aber auch das Wissen verbessern. Die Trefferrate des Systems kann zu zwei Zeitpunkten bestimmt werden. Der erste ergibt sich zum Zeitpunkt t3 zwischen dem Vergleich der ausgewählten Merkmalsinformation in L3 und FB₂₀ (= externes Kriterium). Damit kann überprüft werden, ob in der Lerneinheit auch diejenigen Merkmale enthalten sind, welche im Fragebogen falsch zugeteilt wurden. Die zweite Trefferrate kann über den Vergleich zum Zeitpunkt t4 zwischen den Lerninhalten in L5 und FB₂₀ bestimmt werden. In Abbildung 55 sind die gemittelten Werte, getrennt nach Stichprobe (Adaptiv vs. Standard), aufgeführt.

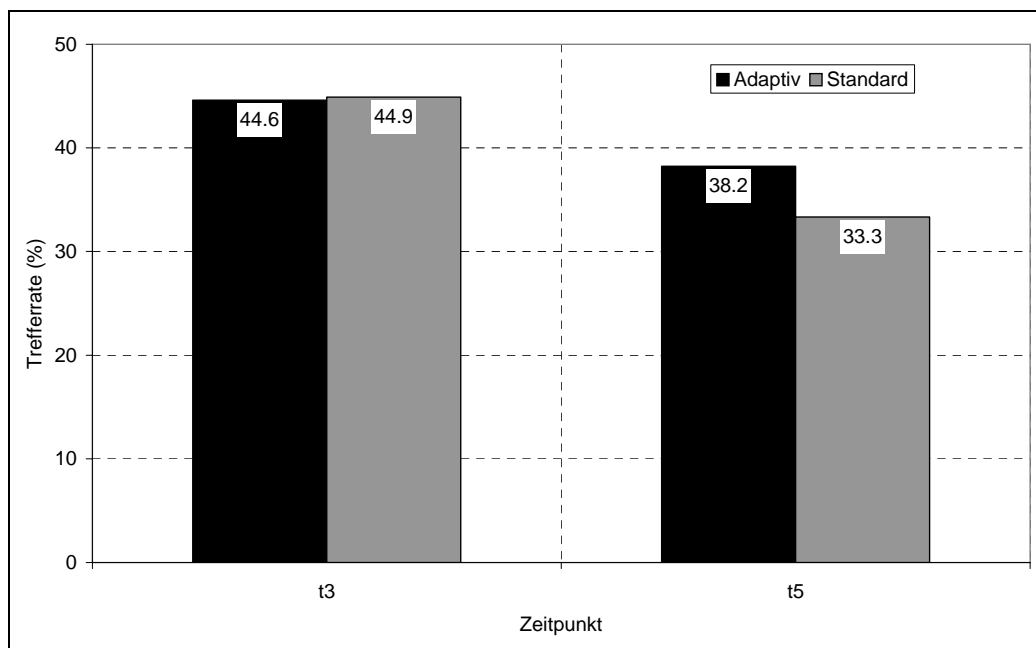


Abbildung 55. Verhältnis von neuer Information zur gesamten Information einer Lerneinheit zum Zeitpunkt t3 und t5.

Die Trefferraten bewegen sich auf einem sehr geringen Niveau. Nicht mal die Hälfte der Merkmalsinformation, welche vom Diagnoseapparat für die kommende Lektion zum Zeitpunkt t3 empfohlen wird, wurde im Fragebogen nicht korrekt beurteilt und dies bei beiden

Stichproben. Zum Zeitpunkt t5 liegt die Trefferrate in der Adaptiv-Gruppe höher als bei der Standard-Gruppe, der Unterschied ist jedoch nicht signifikant ($t(33) = 0.70, p = 0.49$; zweiseitig) und allgemein auf einem noch geringeren Niveau.

Kognitive Karte des Lernalers. Zu zeigen gilt, dass zum einen bei beiden Stichproben sich die Karten aufgrund des Lernsettings dem Normmodell annähern. Zum anderen sollte die Adaptiv-Gruppe einen schnelleren Lernerfolg aufweisen. Abbildung 56 zeigt die gemittelten AvgLoss-Werte der beiden Stichproben über die Zeit.

Auffällig ist, dass sich die Versuchspersonen bei Lernschritt 0 (nach dem Lernen des Vorwissens) immer noch sehr hohe Abweichungswerte zur Normkarte im AvgLoss von über 0.6 produzieren. Als statistischer Vergleichswert kann das Maß von AvgLoss = 0.38 herangezogen werden, welches entsteht, wenn man die berechnete Normkarte basierend auf korrekter und vollständiger Merkmalsinformation mit der berechneten Karte, welche auf der Basis des mit Lücken und Falschwissen definierten Vorwissens basiert, vergleicht. Des Weiteren zeigt sich nur eine geringe Abnahme der Kartenabweichungen über die Lernschritte und erreicht zum Schluss einen Wert von knapp unter 0.6. Die Abnahme der Adaptiv-Gruppe zwischen Lernschritt 0 und 5 ist auf dem 1%-Niveau signifikant ($t(16) = 2.83, p < 0.01$) und erreicht eine mittlere Effektgröße von $d = 0.58$. Die Standard-Gruppe weist ebenfalls eine Abnahme auf dem 1% Niveau auf ($t(17) = 2.84, p < 0.01$) und eine im Vergleich zur Adaptiv-Gruppe marginal geringere Effektgröße von $d = 0.43$ auf. Es ist kein Unterschied zwischen den beiden Gruppen festzustellen. Die beiden Kurven sind praktisch deckungsgleich. Zu keinem Zeitpunkt wird ein Unterschied signifikant.

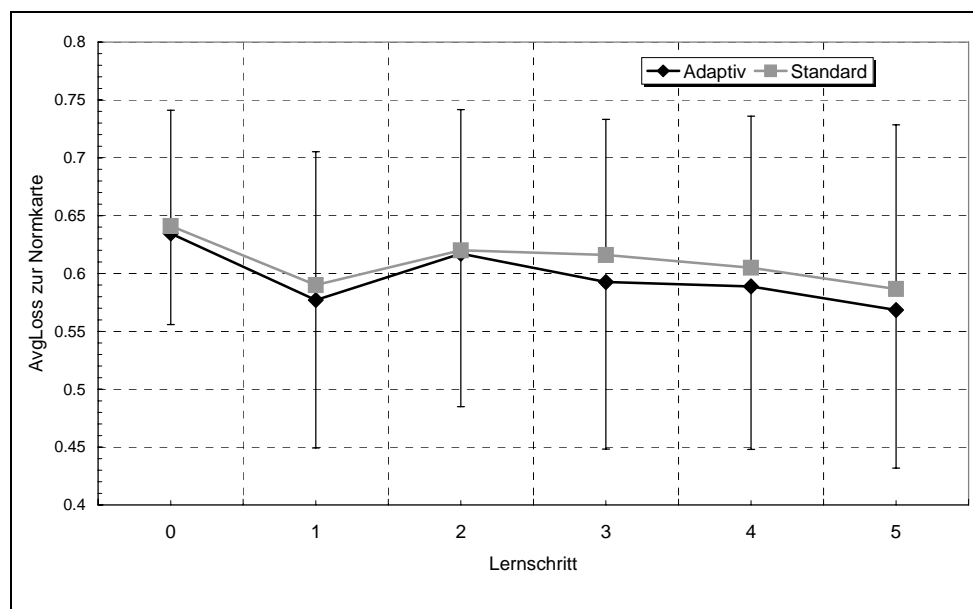


Abbildung 56. Abweichung der Lernerkarten zur Normkarte.

6.4. Diskussion

Entgegen den Erwartungen ließen sich die simulativen Resultate von Daub (2001) experimentell nicht replizieren. Die Trefferraten des Diagnoseapparats sind äußerst tief und unbefriedigend. Bei zwölf möglichen Treffern werden weniger als die Hälfte richtig detektiert. Der Nachweis einer merklichen Kartenverbesserung über das Experiment hinweg blieb ebenfalls aus. Auch wenn statistische Effekte in mittlerer Höhe auftraten, ist der strukturelle Fortschritt auf einer inhaltlichen Ebene ungenügend. Außerdem ist der Absolutbetrag der Kartenabweichung am Ende von t4 viel zu hoch. Ein AvgLoss-Wert von knapp unter 0.6 nach fünfmaligem Lernen ist zwar signifikant vom Zufall verschieden, trotzdem kann nicht konstatiert werden, dass solch eine Karte einem expertennahen Niveau gleichkommt (vgl. für Vergleichswerte Kapitel 2 und 4 in dieser Arbeit). Der postulierte Haupteffekt, dass die Adaptiv-Gruppe aufgrund der Verfahrenslogik einen Lernvorteil zeigen sollte, konnte ebenfalls nicht bestätigt werden.

Es stellt sich nun die Frage nach den Ursachen dieser mangelhaften Resultate. Grundsätzlich bestehen zwei Alternativhypothesen: Zum einen kann vermutet werden, dass das Diagnosekonzept und die didaktische Ausrichtung nicht angemessen ist. Zum anderen könnten Fehlerquellen im Experimentalsetting selbst liegen. Um Ersteres auszuschließen, werden verschiedene Hypothesen überprüft: (1) Verbessern die Lerner überhaupt ihr deklaratives Merkmalswissen? (2) Sind die Kognitiven Karten inhaltlich betrachtet interpretierbar? (3) Sind die Voraussetzungen hinsichtlich des mittleren strukturellen Vorwissens erfüllt?

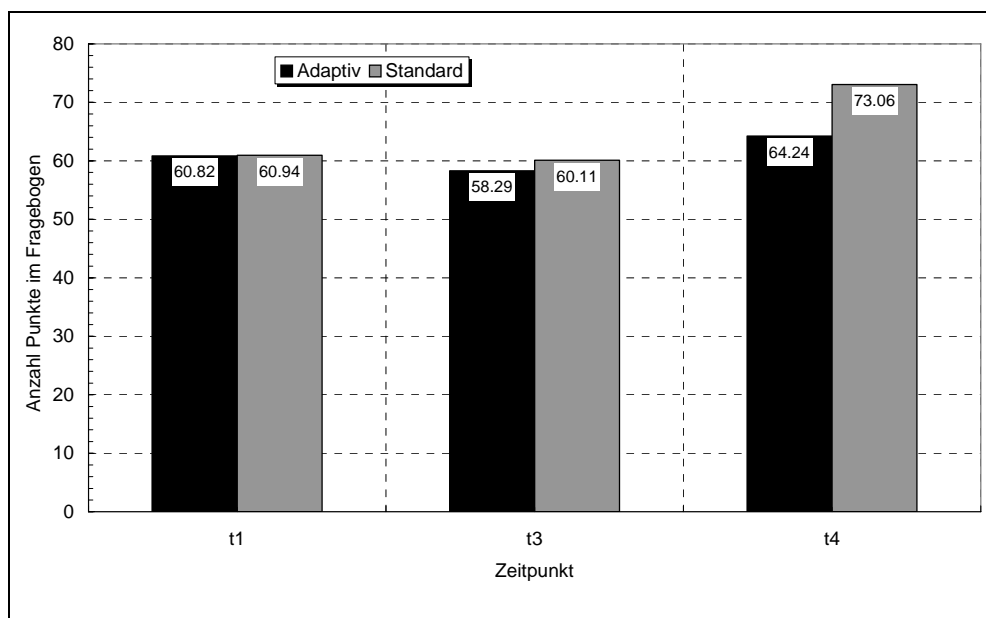


Abbildung 57. Erreichte Punktzahlen im Fragebogen. Maximal mögliche Punkte zum Zeitpunkt t1 = 90; zu den Zeitpunkten t3 und t4 = 120.

Lernfortschritt im Merkmalswissen. Auf struktureller Kartenebene findet kein merklicher Wissenszuwachs statt. Liegt das nun daran, dass die Versuchspersonen gar keine Verbesserung im Merkmalswissen zeigen oder vielmehr am mangelhaften Transfer der Merkmalsin-

formation in adäquate, qualitativ gute Ähnlichkeitseinschätzungen? Die erzielten Punkte in den Fragebögen dienen als externes Überprüfungskriterium. Abbildung 57 zeigt die im Multiple-Choice-Fragebogen erreichten Punktzahlen zum Zeitpunkt t1, t3 und t4.

Zwei Drittel der möglichen Punkte wurden zum Zeitpunkt t1 erreicht. Zwischen Zeitpunkt t3 und t4 ist ein signifikanter Lerngewinn festzustellen (Adaptiv: $t(16) = -4.08, p < 0.001$; Standard: $t(17) = -6.67, p < 0.001$). Dieser Effekt spricht eher dafür, dass die Umsetzung des Merkmalswissens in relationale Beziehungen ungenügend ist, auch wenn absolut betrachtet nur etwas mehr als die Hälfte der möglichen Antworten richtig gelöst wurden. Ein weiteres Faktum muss erwähnt werden, nämlich, dass bei t1 verhältnismäßig mehr gewusst wurde als bei t3 und t4, obwohl bis hierhin bereits viel mehr Lernzeit zur Verfügung stand. Eine plausible Erklärung liegt zum einen in der Anzahl zu beantwortender Fragen. Bei t1 mussten lediglich 45 Merkmalsinformationen beurteilt werden im Gegensatz zu 60 bei t3 und t4, was unbestritten eine schwierigere Aufgabe ist und mehr kognitive Ressourcen verlangt. Außerdem waren bei t1 noch falsche Informationen enthalten, welche ein Umlernen erforderten beziehungsweise bis dato noch nicht in einer Lerneinheit vorkamen und demzufolge nicht korrigiert werden konnten.

Interpretierbarkeit von Kognitiven Karten. Der NMDS-Algorithmus sieht sich mit der schwierigen Aufgabe konfrontiert, (mehrdimensional organisierte) assoziative Daten in einem niedrigdimensionalen Raum zu kartieren. Dies gelingt bei von „Menschenhand“ erzeugten und mit Inkonsistenzen und Ausreißern behafteten Rohdaten nicht perfekt. Der NMDS-Stress ist ein primär technisches Maß, welches Auskunft darüber gibt, wie adäquat das Skalierungsverfahren die Rohwerte in eine Kognitive Karte umzusetzen in der Lage war. Ein hoher Wert deutet darauf hin, dass der iterative Approximationsalgorithmus nicht eine mathematisch optimale Lösung gefunden hat. In technischen Worten ausgedrückt, bedeutet dies, dass das Fehlerminimierungsverfahren in einem lokalen Minimum hängen geblieben ist beziehungsweise aufgrund der Fehlervermeidung eine inhaltlich nicht zu interpretierende Objektpositionierung vollzogen hat, was sich dann in charakteristischen Objektkonfigurationen wie beispielsweise sehr stark geclusterten Lösungen widerspiegeln würde. Die mittleren NMDS-Stress-Werte über alle Diagnosezeitpunkte und beide Stichproben von $M = 0.12$ ($SD = 0.04$) bewegen sich durchgängig unter dem von Gigerenzer (1980) als tolerierbar vorgeschlagenen Niveau bei 20 Objekten und zweidimensionaler NMDS-Lösung von 0.28. Es konnten auch keine, so genannten degenerierten Lösungen beobachtet werden. Damit ist eine wichtige Grundvoraussetzung gegeben, die Kognitiven Karten inhaltlich zu interpretieren und ein technisches „Versagen“ des Verfahrens kann ausgeschlossen werden.

Adäquate Vorwissensstruktur. Abbildung 56 ist zu entnehmen, dass die mittleren AvgLoss-Werte nach dem Lernen und Repetieren des Vorwissens zum Zeitpunkt t2 etwas über 0.6 zu liegen kommen. Potentiell möglich wäre, geht man von sich mathematisch konsequent verhaltenden menschlichen Subjekten bei der Verrechnung von Merkmalswissen in relationale Beziehungen aus, ein Wert von AvgLoss = 0.38 (vgl. Resultate). Dieser Betrag ist sicherlich die untere Grenze des Möglichen, da Personen in ihrem Urteilsverhalten immer eine gewisse Inkonsistenz im Urteilsverhalten aufweisen, welche den minimal möglichen Erwartungswert etwas in die Höhe treiben dürfte. Das fast doppelt so hohe Niveau ist damit aber dennoch

nicht zu erklären. Ein Vergleich der Kartengüte mit der Korrektheit des Merkmalswissens entfällt, da zu diesem Zeitpunkt kein Fragebogen erhoben wurde. Es drängt sich die Interpretation auf, dass die Lerner nicht in der Lage waren, bis zu diesem Zeitpunkt t_2 das Vorwissen vollständig aufzunehmen, geschweige denn eine differenzierte Kognitive Struktur zu bilden. Trotzdem produzieren sie, wie im vorigen Abschnitt beschrieben, zueinander passende Ähnlichkeitsurteile. Daraus lässt sich schließen, dass keine genügende Wissensqualität erreicht wurde. Diese Voraussetzung muss aber zwingend gegeben sein, denn wie wir in Kapitel 4 experimentell und simulativ zeigen konnten, benötigt eine valide Verfahrenslogik einer Wissensdiagnostik über die Prokrustes-Transformation ein bestimmtes Maß an Vorwissen, was gleichbedeutend ist mit einem AvgLoss-Wert von mindestens < 0.6 . Erst dann entwickelt sich eine genügend hohe Stabilität im individuellen Urteilsverhalten, damit der NMDS-Algorithmus die strukturelle Charakteristik der Ausgangsdaten optimal extrahieren kann. Das Wissensdefizit darf im Maximum 40% betragen, um über den ObjLoss einer Prokrustes-Transformation ein Objekt mit genügend hoher Vorhersagegenauigkeit als mangelhaft gewusst zu deklarieren. Bedenkt man außerdem, dass dieser Wert von 40% aus simulativen Daten stammt, wird offensichtlich, dass diese methodische Einschränkung trotz der im Experimentaldesign bewussten Berücksichtigung sich in der empirischen Umsetzung als fatal erwiesen hat. Mit anderen Worten ausgedrückt bedeutet das, dass das Verfahren der Prokrustes-Transformation im Hinblick auf das Erkennen von Wissensdefiziten noch nicht greifen konnte und dass dieser Fehler in der Folge bei der Schätzung des Merkmalswissens über das Kohonen-Netz bestehen blieb.

Das Fazit ist eine mangelhafte Vermittlung des Vorwissens und die dadurch bedingte methodische Unmöglichkeit, valide Wissensschätzungen vorzunehmen. Nicht überraschend ist deshalb auch, dass keine Gruppenunterschiede gefunden werden konnten. Inwiefern dies auch die Performanz der Wissensrepräsentationsform Kohonen-Netz betrifft, kann auf der Basis dieser Daten nicht abschließend beantwortet werden.

Eine detaillierte Auswertung ergab einen interessanten Befund, welcher gleichsam in die Bresche der ungenügenden Voraussetzungserfüllung schlägt und das Verfahren an sich grundsätzlich als tauglich ausweist. Sollte die Methode der NMDS sensitiv auf Lernfortschritte reagieren, müsste nachgewiesen werden können, dass das Lernen von Merkmalswissen die korrespondierende Objektposition qualitativ verbessert. Überprüft werden kann dies anhand des ObjLosses. Das Lernen von Merkmalsinformation zu einem spezifischen Objekt sollte sich in einer nach der Lernphase näheren Position im Vergleich zum Normmodell auswirken. In Abbildung 58 sind die ObjLoss-Werte all derjenigen Länder vor und nach einer Lernsituation aufgezeichnet, zu welchen innerhalb der Lernsituation mindestens eine Merkmalsinformation vermittelt wurde. Alle Unterschiede sind auf dem 0.1%-Niveau signifikant. Die mittleren bis sehr hohen Effektgrößen bewegen sich zwischen $d = 0.63$ (L2) und $d = 1.03$ (L1), welche die Sensitivität des Verfahrens bestätigen. Die Integration von gelerntem Sachwissen in die individuelle kognitive Struktur ist messbar, was eigentlich für das Verfahren spricht, aber dennoch die unerfüllten Voraussetzungen hinsichtlich der Vorwissensstruktur nicht mehr kompensieren kann.

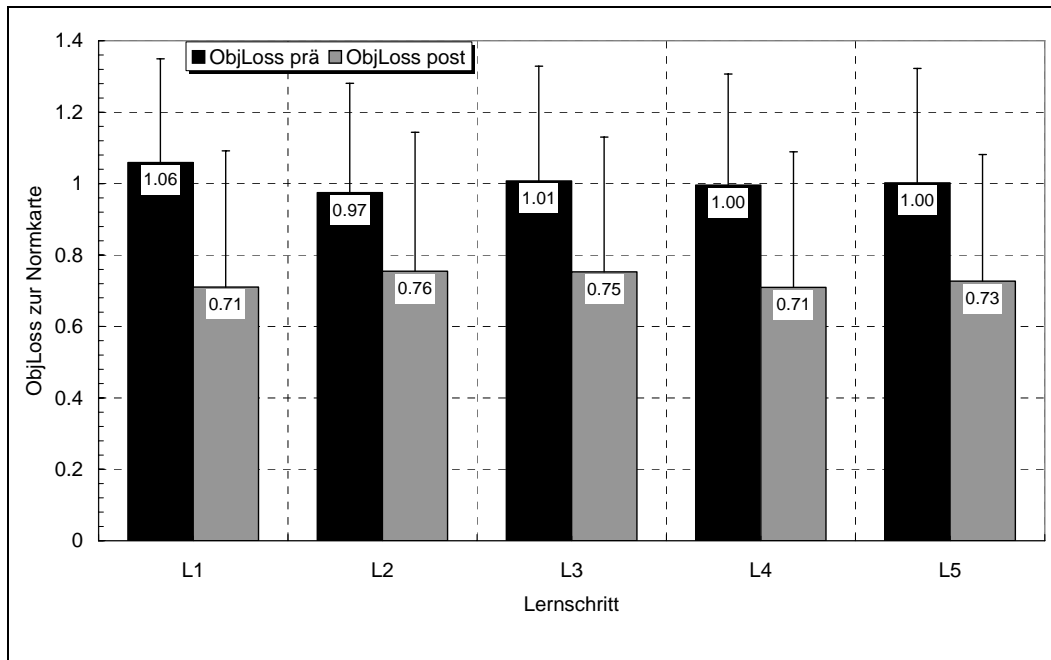


Abbildung 58. Objektabweichung zur Normkarte vor und nach dem Lernen (prä-post) für diejenigen Objekte, welche in der Lerneinheit vorhanden waren (gemittelt über beide Stichproben).

Anwendung im Feld. Die beschriebenen Resultate und Nachauswertungen, welche die mangelhaften Ergebnisse auf das Experimentalsetting zurückführen lassen, würden eigentlich eine erneute Studie erfordern, um der Verfahrenslogik dennoch ihre prinzipielle Tauglichkeit zuzugestehen. Es muss allerdings sehr in Frage gestellt werden, ob der so gewonnene wissenschaftliche Nutzen den dazu erforderlichen Aufwand gerechtfertigt. Rekapitulieren wir, die Bestätigung der Hypothesen und Effekte vorausgesetzt, die für eine praktische Umsetzung des Verfahrens notwendigen Bedingungen für das reibungslose Funktionieren des Diagnoseapparats INTUS: (1) Als erstes müssen die beiden Repräsentationsformen Kognitive Karte und Kohonen-Netz bestmöglich ineinander überführt werden können. Dieser Prozess muss zwar lediglich einmal vorgenommen werden und erfordert lediglich einen vertretbaren technischen Aufwand bei der optimalen Parameterfindung. Nun bestehen aber Inhaltsgebiete (wie auch jenes des hier herangezogenen Länderwissens) nicht nur aus drei definierenden Merkmalsdimensionen, will man ein differenziertes tutorielles System für den Praxiseinsatz konzipieren. Das Kohonen-Netz, welches zwar prinzipiell mehrdimensional trainiert werden kann, aber aufgrund der notwendigen Abgleichung mit den NMDS-Karten lediglich auf zwei Dimensionen beschränkt ist, kann nur bedingt mit mehr als drei Merkmalsdimensionen umgehen (Daub, unpublizierte Daten). Zudem ergeben sich bestimmte Schwierigkeiten beim Training des Kohonen-Netzes, wenn die Merkmalsausprägungen lediglich auf nominalen Skalen organisiert sind, da für die Trainingsvektoren mindestens quasi-intervallskaliertes Niveau vonnöten ist. Benötigt würde somit eine expertisabhängige Ordinalisierung des Wissensgebietes, womit Probleme bei der Interrater-Reliabilität oder der inhaltlichen Verzerrung des Inhaltsbereiches auftreten können. (2) Die zweite Schwierigkeit ergibt sich aufgrund der Einschränkungen das notwendige Maß an Vorwissen betreffend. Kognitive Karten können erst dann als valides Diagnoseinstrumentarium eingesetzt werden, wenn ca. zwei Drittel des Wissensumfangs bereits sicher beherrscht werden. Nur dann erreicht die Prokrustes-Transformation eine

akzeptierbare Diagnoseperformanz (vgl. Kapitel 4). Aber auch dann ergeben sich noch gewisse falsche Vorhersagen. Diese Fehlerquelle, kombiniert mit ebensolchen des Kohonen-Netzes verringert die Diagnoseleistung zu sehr, um die geforderte exakte Wissensschätzung für den praxistauglichen Einsatz eines adaptiven tutoriellen Systems zu gewährleisten.

Resümiert werden muss eine stark zu kontrollierende und komplexe Systemstruktur, die das Programm eher schwerfällig denn flexibel und automatisierbar macht. Trotz aller lerntheoretischen Vorteile und didaktischer Einfachheit bleibt die Kombination von Prokrustes-Transformation und Kohonen-Netz als Diagnosemethode für deklaratives Sachwissen wohl auf den Einsatz für wissenschaftliche Laborstudien beschränkt, innerhalb dessen aber unbestritten Platz für grundlagenorientierte Fragestellungen besteht.

Ausblick. Simulationen ermöglichen es demgegenüber, mit verhältnismäßig geringem zeitlichem Aufwand, große Datenmengen zu generieren. Die immer noch ungeklärte Frage, was Lernraten und Vergessensraten für einen Einfluss auf den Lernprozess, kompetitiv verglichen mit einem „klassischen“ Vorgehen, haben, soll darum in einer Erweiterung der Simulation von Daub (2001) beantwortet werden. Im Folgenden wird nun das Konzept dafür beschrieben, welches, so weit möglich, grundlegende Erkenntnisse aus der Lern- und Gedächtnisforschung umzusetzen versucht. Zu diesem Zweck soll vorerst eine kurze Replik wichtiger Erkenntnisse aus der Gedächtnisforschung umrissen werden, welche die Konzeptualisierung der Simulation untermauern. Das Ziel dieses Überblicks soll sein, wie Lernen und Vergessen simuliert werden könnte.

Seit Ebbinghaus (1971) ist bekannt, dass Lernen und Erinnern einer diskontinuierlichen kognitiven Leistung gleichkommt. Die Vergessenskurve dürfte wohl die bekannteste Erkenntnis seiner anhand des Lernens von Wortlisten gefundenen Charakteristik des Gedächtnisses sein. Sie sagt aus, dass kurz nach dem Lernen ein großer Abfall der Erinnerungsleistung stattfindet im Gegensatz zu geringen Verlusten von sich bereits länger im Gedächtnis befindlichen Materials. Auch bekannt sind seine Resultate zur Lernkurve, welche besagen, dass der größte Lernfortschritt zu Beginn des Wissenserwerbs passiert. Und, je mehr man bereits weiß, desto mehr Lernaufwand muss betrieben werden, um neue Information aufzunehmen. Die Bestimmung solcher Lern- und Vergessensfunktionen wären simulativ leicht umsetzbar. Nicht unerwähnt bleiben darf aber die Kritik an Ebbinghaus' untersuchtem Material (z.B. Baddeley, 1979, bezugnehmend auf Bartlett, 1932). Angefochten wurde vor allem die mechanistische Sichtweise beim Lernen von sinnlosen Wortlisten, welche im Vergleich zu dem „in der Natur“ auftretendem realen Lernen eher einem Artefakt gleichkommt. Dass Wissen nur selten in der Art eines „Nürnberger Trichters“ aufgenommen wird, wurde beispielsweise auch von Ausubel, Novak & Hanesian (1980/81) oder Bruner (1974) in ihren Arbeiten zur Unterrichtsforschung propagiert. Es muss somit unterschieden werden zwischen mechanischem und sinnvollem Lernen, wobei letzteres die inhaltliche (assimilative) Verknüpfung von neuem Wissen mit bestehendem Alten bedeutet und zu einer größeren Verarbeitungstiefe führt (also auch besser behalten wird). In der Simulation und im Experiment ging es zudem nicht darum, reine serielle Wortlisten auswendig zu lernen, sondern semantisch bedeutsame Paarassoziationen zu bilden. Das heißt, dass die Ebbinghaus'schen Lern- und Vergessenskurven wohl nicht durch eine generalisierbar fixe, rein von der Anzahl und der Latenzzeit zwischen Lernen und Abruf abhängigen

Funktion beschrieben werden können. Nichtsdestotrotz kann angenommen werden, dass die prinzipielle Aussage von Ebbinghaus, dass Neugelerntes schneller vergessen wird als schon länger Gespeichertes und dass auch nicht alles Material auf einmal aufgenommen werden kann, seine Berechtigung hat. Lediglich die mathematisch quantifizierbare Charakteristik der Lern- und Vergessenskurven wird sich unterscheiden, je nachdem mit welchem Lernmaterial man sich auseinandersetzt und wie das Vorwissen strukturiert ist.

Einen zweiten Forschungszweig bilden Versuche, das Langzeitgedächtnis durch mathematische Modelle zu beschreiben und unterscheiden sich von der vorher kurz angesprochenen Auffassung von Verarbeitungstiefen der individuellen Gedächtnisrepräsentationen. Vor allem die globalen Gedächtnismodelle sind prominente Beispiele, nach deren zentralen Annahme die Erinnerung an ein bestimmtes Ereignis immer durch alle im Gedächtnis repräsentierten Inhalte bestimmt ist (vgl. für einen kurzen Überblick über globale Gedächtnismodelle: Buchner & Brandt, 2002). Beispielhaft sei das Modell MINERVA2 (Hintzmann, 1986, 1988) genannt. Neben strukturellen Aspekten des Gedächtnisses wie bspw. die Elemente (Objekte) mit ihren Attributen (Merkmale) sind in MINERVA2 Lern-, Vergessens- und Abrufprozesse definiert. Der Lern- und der Vergessensparameter bestimmen, mit welcher Wahrscheinlichkeit ein Attribut gelernt bzw. vergessen wird. Dieses einfach strukturierte Modell kann eine Vielzahl an empirischen Ergebnissen zur Gedächtnisforschung erklären (Buchner & Brandt, 2002). Allerdings können auch daraus nicht feste Lern- und Vergessensparameter abgeleitet werden, wohl aus dem Grund, dass solche Gedächtnismodelle eher dazu dienen, empirische Resultate nachträglich über Simulationen nachzuzeichnen.

Nicht unerwähnt bleiben dürfen Erkenntnisse aus der Erforschung von Gedächtnistäuschungen, auch Fehlerinnerungen genannt. Sie treten auf, da das Gedächtnis bzw. die Erinnerung keinem passiven, sondern ein aktiv konstruierenden Kräften unterliegt. Zwei sich konkurrierende Hypothesen zur Erklärung von Täuschungen wie bspw. der „Misleading Postevent Information“ oder des „Hindsight-Bias“ machen die Runde. Zum einen die so genannte Integrationshypothese (Loftus, Miller & Bruns, 1978) und zum anderen die Koexistenzhypothese (McCloskey & Zaragoza (1985)¹². Grundsätzlich beschreiben die beiden Hypothesen nichts anderes, als dass gelerntes (korrektes) Wissen, welches durch nachträgliche (externe) Konfrontation mit neuer Information verfälscht wird. Die Integrationshypothese besagt vereinfacht, dass neue (konkurrierende Information) in die bestehende Gedächtnisspur integriert wird. Dies würde bedeuten, dass eine Art „Mittelwertsmodell“ des Wissens gebildet wird. Die Erinnerung konstituiert sich demzufolge über eine Gedächtnisrepräsentation, welche sich „in der Mitte“ zwischen ursprünglicher falscher und neu gelernter korrekter Information befindet. Im Gegensatz dazu die Koexistenzhypothese, nach der beide Gedächtnisspuren nebeneinander ko-existieren, ohne sich gegenseitig zu beeinflussen. Die Stärke der Gedächtnisspuren kann jedoch unterschiedlich sein und führt dann zu einem Erinnerungsfehler, wenn sich die Person

¹² Hell (1993) gibt einen Überblick über Forschungsergebnisse, die zum einen empirische Ergebnisse erklären können, zum anderen aber auch nur für die eine oder die andere Variante sprechen. Da sich bislang kein Modell endgültig durchgesetzt hat, werden in der vorliegenden Arbeit beide Hypothesen als Grundlage für die simulative Umsetzung herangezogen. Auch Hell (1993) schlägt eine Mischform dieser beiden Hypothesen vor.

nur noch an die Fehlinformation erinnert (für eine kritische Gegenüberstellung der beiden Hypothesen vgl. Hell, 1993). Dies bedeutet für die hier zu konzipierende Simulation von Lernen und Vergessen, dass das Falschwissen des Künstlichen Lernalers, welches über den Lernprozess korrigiert werden soll, nicht in jedem Fall zu korrektem Wissen führt, sondern tendenziell auch falsch bleiben kann. Mit anderen Worten, die Information wird nicht gelernt.

Die kurz angeschnittenen Forschungsergebnisse und theoretischen Modellannahmen lassen die Komplexität des Sachverhalts erahnen. Es können nicht allgemeine und quantifizierbare Verlaufskonstanten das Lernen und Vergessen betreffend bestimmt werden. Für die angestrebte eigene Simulation wird man somit nicht umhin kommen, ebenfalls verschiedene Möglichkeiten und Parametereinstellungen durch zu testen. Es werden darum sehr einfache Varianten von Lernen und Vergessen simuliert. Der Rest der Simulation bleibt exakt gleich, es besteht jetzt lediglich in der Lernphase die Möglichkeit, unterschiedliche Kombinationen von Lern- und Vergessensraten durchzuspielen. Eine die obigen Ausführungen stark vereinfachende Variante wird vorgestellt:

„Dummy“-Variante (Abbildung 59): Die Lernrate auf der linken Seite folgt einem Alles-oder-Nichts-Prinzip (vgl. Koexistenzhypothese). Entweder wird das Element korrekt gelernt oder verbleibt auf dem alten Wert. Man gibt an, wie viele der in der Lektion enthaltenen Merkmale perfekt gelernt werden (in MINERVA2 würde dies eine Lernwahrscheinlichkeit von 100% bzw. 0% bedeuten, was zwar einer konstruierten Sichtweise gleichkommt, vom Modell her aber prinzipiell zulässig ist). Die nicht gelernten Merkmale werden zufällig bestimmt. Die Vergessensrate wird durch zwei Parameter definiert. Zum einen die Anzahl derjenigen Einheiten (rechts oben in Abbildung 59), welche gelernt worden sind, nun aber wieder vergessen werden (bei MINERVA2: Vergessenswahrscheinlichkeit von 100% bzw. 0%). Zum anderen der Faktor a , welcher angibt, um wie viel Prozent das Merkmal wieder in die Richtung des ursprünglichen Wertes „verschoben“ wird (vgl. Integrationshypothese).

The screenshot shows a software interface titled "Dummy". It contains a checkbox that is currently unchecked. Below the checkbox, there are two input fields. The first field is labeled "Wieviele Einheiten aus Lerneinheit Perfekt lernen" and contains the number "0". The second field is labeled "Wieviele neugelernte Einheiten aus Lerneinheit um Faktor a verändern" and also contains the number "0". Below the second field, there is a label "Faktor a" and another input field containing the number "0".

Abbildung 59. Screenshot der Eingabeoberfläche.

Was in dieser Simulation natürlich noch fehlt, ist der Einbezug von Ebbinghaus' Erkenntnissen, nämlich dass mit fortschreitender Lerndauer mehr Wiederholungen notwendig sind, um einen Lernfortschritt zu erzielen und dass auch Wissen, welches vor einiger Zeit erworben worden war, auch noch zu einem gewissen Grad vergessen wird. Umsetzbar wäre dies dadurch, dass bereits gelernte und auch wieder vergessene Merkmale eine immer kleiner werdende Wahrscheinlichkeit besitzen, im Wissen des Künstlichen Lernalers korrekt abgespeichert zu werden. Des Weiteren müssten einmal gelernte und bis zu einem gewissen Grad vergessene Merkmale nach jedem Lernschritt, innerhalb dessen diese Elemente nicht vorkommen, kontinuierlich noch weiter vom korrekten Wert abweichen.

Unbestritten ist die gewählte Umsetzung von Erkenntnissen aus der Gedächtnisforschung in der beschriebenen Form ein sehr vereinfachtes, der Vielschichtigkeit der Sachlage wohl nicht vollumfänglich gerecht werdendes Vorgehen und Kritik ist sicherlich berechtigt. Jedoch ist das Ziel dieser Simulation nicht, kognitive Prozesse in voller Gültigkeit nachzubilden. Vielmehr geht es darum den postulierten Adaptivitätsvorteil im Lernfortschritt hinsichtlich der Wissensvermittlung zu bestätigen. Im Prinzip genügt es also, wegzukommen von der Annahme einer perfekten Lernrate und einer nicht vorhandenen Vergessensrate wie dies noch bei Daub (2001) implementiert war. Es wird interessant sein, welche Kombination der Parametereinstellungen den größten Unterschied zwischen adaptivem und standardisiertem Lernprozess bringt. Daraus und unter Einbezug von Erkenntnissen aus der Unterrichtsforschung können Empfehlungen bezüglich des optimalen Lernumfangs, der Latenzzeit zwischen den Lektionen o.ä. generiert werden und somit wäre ein weiterer Beitrag zur Konstruktionsanleitung für automatisierbare tutorielle Systeme geleistet.

6.5. Literatur

- Ausubel, D. P., Novak, J. D. & Hanesian, H. (1980/81). *Psychologie des Unterrichts*. 2 Bde. (2. Aufl.). Weinheim: Beltz.
- Baddeley, A. D. (1979). *Die Psychologie des Gedächtnisses*. Stuttgart: Klett – Cotta.
- Bartlett, F. C. (1932). *Remembering*. Cambridge: Cambridge University Press.
- Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling – Theory and applications*. New York: Springer.
- Bruner, J. S. (Hrsg.). (1974). *Lernen, Motivation und Curriculum*. Frankfurt a.M.: Athenäum.
- Buchner, A. & Brandt, M. (2002). Gedächtniskonzeptionen und Wissensrepräsentationen. In J. Müsseler und W. Prinz (Hrsg.), *Allgemeine Psychologie* (S. 492-543). Heidelberg: Spektrum Akademischer Verlag.
- Crowder, N. A. (1959). Automating tutoring by means of intrinsic programming. In E. Galanter (Ed.), *Automatic teaching. The state of the art* (pp. 109-116). New York: John Wiley and Sons, Inc.
- Daub, S. (2001). *Similarity Simulation – Wie man den Code des Globaureils knackt*. Dissertationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Der Fischer Weltalmanach (2002). *Staaten. Der Fischer Weltalmanach – Online Datenbank*. Aufgerufen 3. Juli, 2002 unter http://www.weltalmanach.de/staat/staat_liste.html
- Ebbinghaus, H. (1971). *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie*. Darmstadt: Wissenschaftliche Buchgesellschaft (Nachdruck der Ausgabe von 1885).
- Gower, J. C. & Dijksterhuis, G. B. (2004). *Procrustes Problems*. New York: Oxford University Press.
- Hell, W. (1993). Gedächtnistäuschungen. In W. Hell, K. Fiedler & G. Gigerenzer (Hrsg.), *Kognitive Täuschungen: Fehl-Leistungen und Mechanismen des Urteilens, Denkens und Erinnerns* (S. 13-38). Heidelberg: Spektrum Akademischer Verlag.
- Hintzmann, D. L. (1986). „Schema abstraction“ in a multiple trace memory model. *Psychological Review*, 93, 411-428.
- Hintzmann, D. L. (1988). Judgments of frequency and recognition memory in a multiple trace model. *Psychological Review*, 41, 109-139.
- Klauer, K. J. (1989). Allgemeine oder bereichsspezifische Transfereffekte eines Denktrainings. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 21, 185-200.
- Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.
- Kohonen, T. (1984). *Self organization and associative memory* (2nd ed.). Berlin: Springer.

- Kohonen, T. (1995). *Self-organizing maps*. Berlin: Springer.
- Läge, D. (2001). *Ähnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Läge, D., Daub, S. Bosia, L., Jäger, C. & Ryf, S. (2005). *Die Behandlung ausreißer-behafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag*. AKZ-Forschungsbericht Nr. 21. Zürich: Angewandte Kognitionspsychologie.
- Leutner, D. (1992a). *Adaptive Lernsysteme. Instruktionspsychologische Grundlagen und experimentelle Analysen*. Weinheim: Psychologie Verlags Union.
- Leutner, D. (1992b). Das Testlängendilemma in der lernprozess-begleitenden Wissensdiagnostik. *Zeitschrift für Pädagogische Psychologie*, 24, 232-248.
- Loftus, E. F., Miller, D. G. & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Verbal Learning and Verbal Behavior*, 13, 585-589.
- Marx, W., & Hejj, A. (1989). *Subjektive Strukturen*. Göttingen: Hogrefe.
- Marx, W., & Läge, D. (1995). *Der ideologische Ring*. Göttingen: Hogrefe.
- McCloskey, M. & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General*, 114, 1-16.
- Pressey, S. L. (1926). A simple apparatus which gives tests and scores and teaches. In (reprint) A. A. Lumsdaine & R. Glaser (1960), *Teaching machines and programmed learning* (pp. 35-41). Washington D.C.: National Education Association of the United States.
- Skinner, B. F. (1958). Teaching machines. In (reprint) A. A. Lumsdaine & R. Glaser (1960), *Teaching machines and programmed learning* (pp. 137-158). Washington D.C.: National Education Association of the United States.

7.

A Structure-Based Expert Model of the ICD-10 Mental Disorders

Abstract

Many authors have highlighted that a classificatory system such as the ICD-10 or the DSM-IV should attempt to integrate categorical as well as dimensional aspects rather than relying only on a categorical distinction. In the current study, a method is presented that allows both perspectives to be considered. Based on their clinical experience, twenty therapists were asked to rate a selection of ICD-10 mental disorders in terms of their similarity. The resulting data was processed by nonmetric multidimensional scaling (NMDS) on individual basis as well as mean basis and a cluster analysis. From a categorical point of view, the resulting mean cognitive map and the cluster analytic groups showed a good concordance with the ICD-10 F groups. Additionally, analogies to the critical observations with regard to the classification systems could be observed. From the dimensional point of view, there were two dimensions found (cognition and psychosis) which showed high correlations in the mean map. In conclusion, the method which is presented in this study offers an opportunity to account for dimensional as well as categorical aspects of classification by exploring the structure of the ICD-10 mental disorders directly from the point of view of therapists and thus creating the basis for an expert model.

Keywords: ICD-10, mental disorder, expert model, cognitive structure, cognitive map, similarity judgment, nonmetric multidimensional scaling (NMDS)

7.1. Introduction

One of the main challenges of a classification system in psychopathology such as the ICD-10 Classification of Mental Disorders (World Health Organization, 1992) or the DSM-IV (American Psychiatric Association, 1994) that has repeatedly been highlighted is the difficulty of determining clear-cut boundaries between different illness entities in general (Clark, Watson & Reynolds, 1995; Dittmann, 1996; Widiger & Clark, 2000) and between various diagnostic categories in particular. For instance, Maser & Cloninger (1990) argued in “Comorbidity of Mood and Anxiety Disorders” that “In other words, psychopathology does not appear to be comprised of discrete, mutually exclusive disorders, as Kraepelin originally proposed”. Another prominent diagnostic category which has recently gained greater attention is the category of personality disorders. In this diagnostic area, too, the suggestions for an adaptation of the classification tend away from a pure categorical model towards a combination of a categorical and a dimensional model (Mundt, 2002; Oldham & Skodol, 2000; Skodol, Gunderson, Pfohl, Widiger, Livesley & Siever, 2002). Indeed, the debate of whether a categorical or a dimensional system is to be favored, which generally results in suggestions for a combination of the two, has been going on for many years (Blashfield, 1990; Clark, Watson & Reynolds, 1995). The aim of the present study is to integrate both perspectives by employing multivariate structure analyses, especially nonmetric multidimensional scaling (NMDS). These methods allow the ICD-10 classification of mental disorders to be looked at from the point of view of

therapists (i.e. psychiatrists and psychotherapists) with the aim of establishing the basis for an expert model of mental disorders.

7.2. *Methods*

Subjects. The sample consisted of 20 practicing therapists (8 female and 12 male) in Zurich, Switzerland. Eight of the therapists hold a medical degree, eleven hold a psychological degree and one holds both a medical and a psychological degree. Their mean clinical experience was $M = 12.75$ ($SD = 9.8$) years (min = 1.5 years, max = 30 years). In their clinical practice they apply cognitive behavioural $N = 3$, psychodynamic -/analytic $N = 2$, integrative/eclectic $N = 13$, and other $N = 2$ therapeutic approaches.

Tasks. The experiment consisted of two tasks: First the subjects had to rate 210 presented pairs of 21 disorders (e.g. paranoid schizophrenia and dysthymia) based on their similarity on a scale from 1 = minimal similarity to 9 = maximal similarity (similarity judgments: Coxon, 1982; Daub, 2001; Laege, 2001; Smith, Shoben & Rips, 1974). Subjects were specifically instructed not to attempt to reproduce the categorization of the ICD-10, but to make judgments according to their clinical experience. Second, subjects were required to separate the same selection of disorders into groups based on their similarity (parallel sorting task: Laege, 2001) in reference to (a) phenomenology and (b) aetiology. Additionally, after completing the two tasks participants answered several questions regarding their education and their professional experience, and completed ratings concerning a) the subjective importance of the ICD-10 in their clinical everyday life (1 = not important at all to 9 = very important), b) their subjective expertise in the ICD-10 (1 = nearly no knowledge to 9 = very good knowledge), c) the subjective positive attitude towards the ICD-10 (1 = not very positive to 9 = very positive), and d) their potential preference of another classification system. The sequence of the presented pairs of disorders in the first task, the two sorting criteria in the second task, and the questions regarding the ICD-10 were randomized. All tasks were presented on a computer screen with the INTUS[®] data collection software (Ryf, Daub & Laege, 2003) (except for the four ratings, which were given using paper and pencil).

Selected Disorders. The 21 disorders which had to be rated were selected based on (a) representativeness and variability with regard to the field of knowledge (i.e. the psychopathological disorders as defined by the ICD-10) and (b) clinical prevalence (based on a common ranking of three studies with regard to clinical prevalence of mental disorder diagnoses (Lange, Munk-Jorgensen, Bertelsen, Schurmann, Michels, Malchow & Dilling, 2002; Mussigbrodt, Michels, Malchow, Dilling, Munk-Jorgensen & Bertelsen, 2000; Neuenschwander, Meyer & Hell, 1999)). Mental retardation (F7), disorders of psychological development (F8), and behavioural and emotional disorders with onset usually occurring in childhood and adolescence (F9) were excluded from this experiment. Ultimately, three disorders from each category from F0 to F6 were chosen, which added up to the total of the following 21 disorders: Dementia in Alzheimer's disease with early onset (F00.1); organic delusional (schizophrenia-like) disorder (F06.2); organic personality disorder (F07.0); mental and behavioural disorder due to the use of alcohol, withdrawal state (F10.3); mental and behavioural disorder due to the use of

cocaine, psychotic disorder (F14.5); mental and behavioural disorder due to multiple drug use and use of other psychoactive substances, dependence syndrome (F19.2); paranoid schizophrenia (F20.0); hebephrenic schizophrenia (F20.1); schizoaffective disorder, manic type (F25.0); bipolar affective disorder, current episode manic without psychotic symptoms (F31.1); severe depressive episode without psychotic symptoms (F32.2); dysthymia (F34.1); agoraphobia (F40.0); mixed anxiety and depressive disorder (F41.2); adjustment disorder (F43.2); anorexia nervosa (F50.0); bulimia nervosa (F50.2); nonorganic insomnia (F51.0); paranoid personality disorder (F60.0); emotionally unstable personality disorder (F60.3) and enduring personality change after catastrophic experience (F62.0).

Analysis. The data of the similarity judgments and of the parallel sorting task were translated into matrices and then analyzed by robust nonmetric multidimensional scaling (NMDS by means of the ROBUSCAL algorithm; Laege, Daub, Bosia, Jaeger & Ryf, 2005) and a hierarchical cluster analytic method (average model; Bortz, 1999; Johnson, 1967; Wiedenbeck & Zuell, 2001). The NMDS analysis resulted in cognitive maps which were calculated based on individual matrices as well as on the mean matrix of the similarity judgments. The cluster groups resulting from the cluster analysis were calculated based on the mean similarity matrix as well as on the distance matrix, which was extracted from the cognitive map calculated based on the mean similarity matrix. The cognitive maps depict the cognitive relational structure of the subjects in the form of a Euclidian space, where a small distance between the objects (i.e. disorders) corresponds to a high similarity and vice versa. This Euclidian metrics significantly differs from a city-block metrics of a factor analysis. In a factor analysis, the distance between two items can only be described as an arithmetic operation of the values of these points on the corresponding factors/ dimensions. In contrast, the Euclidian space constructed by the NMDS allows the relational position of the items to be directly described and interpreted. Additionally, the dimensions in this space do not need to be orthogonal to each other, as is compulsory for the dimensions/ factors of a factor analysis. The individual maps were compared to each other as well as to the map calculated based on mean matrix by means of a multiple procrustean transformation (Borg & Groenen, 1997; Coxon, 1982; Hurley & Cattell, 1962) (in this study the specific PROPERSCAL method was employed; Laege, 2001). These comparisons resulted in deviation measures (average losses; Laege, 2001) between the maps. Based on the deviations of the individual maps from each other, a two-dimensional map was calculated. This map represents the relational position of the therapists based on their individual cognitive structures (i.e. their cognitive maps). In the resulting map the data with regard to professional experience, therapeutic approach, academic training and attitude towards the ICD-10 was examined for structure explicative value (meaningful clusters and multiple regressions; Property Fitting; Laege, 2001).

In addition, we conducted a content analysis (comparative, semantic -semantic category analysis) (Merten, 1995) of the ICD-10 clinical descriptions and diagnostic guidelines for each of the selected disorders. In a first step, three dimensions/ categories (Cognition, Psychosis and Affect) were defined, derived from the AMDP System (Arbeitsgemeinschaft für Methodik und Dokumentation in der Psychiatrie, 2000). In a second step, the frequency of terms occurring in the text associated with each of the dimensions was determined. This frequency served

as an indicator of the emphasis of each disorder on the corresponding dimension. In a third step, these indicators were fitted into the two dimensional cognitive map (based on the mean similarity matrix) by Property Fitting. All statistical analyses were performed with Microsoft[®] Excel[®] (Microsoft Corporation, 2002), SPSS[®] (SPSS Inc., 2001), INTUS[®] (Ryf, Daub & Laege, 2003) and ASSO[®] (Schnyder, Daub, Jaeger, Ryf, Christen & Laege, 2003).

Evaluation. In a second meeting, the cognitive maps and the cluster groups were presented to those therapists who could be reached again for the evaluation (13/20). They were asked to interpret the cognitive maps based on individual and mean matrices and the cluster groups based on similarity and distance matrices with regard to clinical meaning and their subjective preference. The therapists were blinded to the calculation basis of the cognitive maps and the cluster groups, and the sequence of the presentation was randomized.

7.3. Results

The core result of this investigation is represented by the cognitive map based on the mean similarity matrix in Figure 1. This map depicts the therapists' mean cognitive relational structure of the selected mental disorders. As will be demonstrated, it enables categorical as well as dimensional structural aspects to be interpreted. Before looking at the map with regard to content, it is necessary to explain an internal quality measure of the ROBUSCAL algorithm: The standardized stress value (Laege, Daub, Bosia, Jaeger & Ryf, 2005) is a measure of how well the algorithm was able to translate the similarity judgments into an n-dimensional map. Thus it is also an indicator of the level of consistency of the similarity judgments given by the participants throughout the experiment, and consequently provides a marker of the interpretability of the cognitive map. According to the literature (Borg & Groenen, 1997; Gigerenzer, 1981) the stress value of this mean map of 0.17 as well as the stress values of the individual two-dimensional cognitive maps ($M = 0.16$, $SD = 0.02$) are more than acceptable, and thus interpretable.

As mentioned above it is possible to look at the structure of the cognitive map in Figure 1 from a categorical as well as from a dimensional perspective: The plotted delineations comprise those disorders that are subsumed in the same F category according to the ICD-10, and thereby represent a traditional categorical point of view. As described in the analysis section, the cognitive maps allow the researcher to observe the Euclidian distances between the objects (i.e. disorders). Consequently it is also possible to interpret the relative positions of the disorders in reference to each other, which opens up a dimensional perspective. The examination of the cognitive maps for explicative dimensions was carried out by means of multiple regressions (Property Fitting). The values used in the Property Fitting resulted from the content analysis described in the analysis section. The two arrows in Figure 1, which are labelled "psychosis" and "cognition", depict the orientation of the corresponding dimensions in the map. The emphasis in the ICD-10 of a given disorder on the corresponding dimension increases with its orthogonal projection on the dimension in the direction of the arrow. The following disorders showed a high correlation ($r = 0.80$, $R^2 = 0.64$) on (a) the cognition dimension (Dementia in Alzheimer's disease with early onset (F00.1); organic delusional (schizophrenia

-like) disorder (F06.2); organic personality disorder (F07.0); mental and behavioural disorder due to multiple drug use and use of other psychoactive substances, dependence syndrome (F19.2); paranoid schizophrenia (F20.0); hebephrenic schizophrenia (F20.1); bipolar affective disorder, current episode manic without psychotic symptoms (F31.1); severe depressive episode without psychotic symptoms (F32.2); agoraphobia (F40.0); mixed anxiety and depressive disorder (F41.2); nonorganic insomnia (F51.0); paranoid personality disorder (F60.0) and emotionally unstable personality disorder (F60.3)), and a fairly high correlation ($r = 0.61$, $R^2 = 0.37$) on (b) the psychosis dimension (organic delusional (schizophrenia -like) disorder (F06.2); organic personality disorder (F07.0); mental and behavioural disorder due to the use of cocaine, psychotic disorder (F14.5); mental and behavioural disorder due to multiple drug use and use of other psychoactive substances, dependence syndrome (F19.2); paranoid schizophrenia (F20.0); hebephrenic schizophrenia (F20.1); schizoaffective disorder, manic type (F25.0) and paranoid personality disorder (F60.0)). The selected disorders showed only a low correlation on the affect dimension ($r = 0.33$, $R^2 = 0.11$), which is consequently not presented in the map.

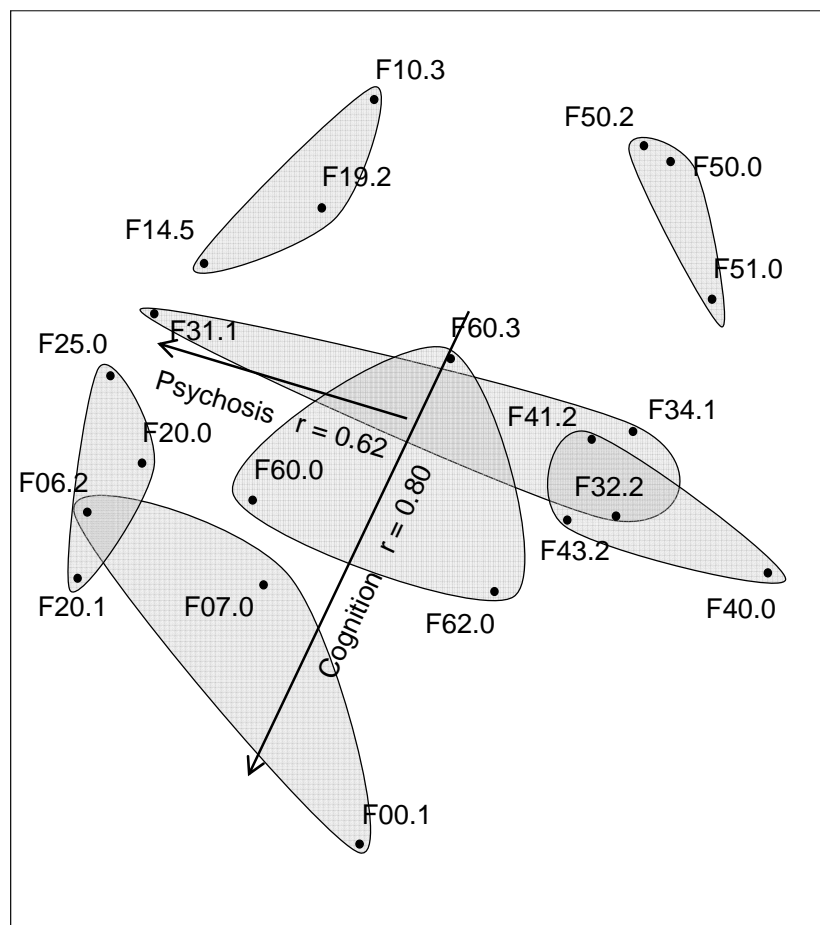


Figure 1. Two-dimensional cognitive map based on the mean similarity matrix (The plotted clusters depict the ICD-10 F categories; Dimensions: Cognition, $r = 0.80$, $R^2 = 0.64$; Psychosis, $r = 0.61$, $R^2 = 0.37$).

In addition to the possibility of plotting the delineations of the disorders in the cognitive maps, which is described above, there is a further method for looking at the similarity matri-

ces from a categorical point of view: The cluster analysis subsumes homogeneous (i.e. similar) objects in clusters in such a way that objects of different clusters are as heterogeneous as possible (Bortz, 1999). The cluster analysis of the mean similarity matrix resulted in six groups of similar disorders (Table 1). The detailed procedure has been described elsewhere (Bortz, 1999; Wiedenbeck & Zuell, 2001). This reanalysis of the similarity matrices by cluster analysis allows the categories found by the NMDS analysis in the cognitive maps to be confirmed or refuted.

Table 1. Groups resulting from the cluster analysis based on the mean similarity matrix.

F40.0 Agoraphobia
F62.0 Enduring personality change after catastrophic experience F43.2 adjustment disorder F41.2 mixed anxiety and depressive disorder F34.1 dysthymia F32.2 severe depressive episode without psychotic symptoms F51.0 nonorganic insomnia
F50.0 anorexia nervosa F50.2 bulimia nervosa F60.3 emotionally unstable personality disorder
F31.1 bipolar affective disorder, current episode manic without psychotic symptoms F25.0 schizoaffective disorder, manic type F20.1 hebephrenic schizophrenia F06.2 organic delusional (schizophrenia like) disorder F20.0 paranoid schizophrenia F60.0 paranoid personality disorder
F10.3 mental and behavioral disorder due to the use of alcohol, withdrawal state F14.5 mental and behavioral disorder due to the use of cocaine, psychotic disorder F19.2 mental and beh. dis. due to mult. drug use and use of other psychoactive subst., depend. Syndr.
F00.1 Dementia in Alzheimer's disease with early onset F07.0 organic personality disorder

Although we have so far only been looking at the analyses of the mean results, the NMDS also allows cognitive maps to be calculated and interpreted on an individual level. This allows an examination of whether the mental structures of the therapists expressed as cognitive maps rather converge or diverge from each other. Figure 2 shows the result of the comparison (Procrustean transformation) of the individual cognitive maps of two therapists chosen as an example. The connecting lines between the disorders represent their deviation of their spatial positions in the cognitive maps of the corresponding therapists. The mean average loss of the comparisons of the individual maps to each other ($M = 0.59$, $SD = 0.09$) as well as the average losses of the comparisons of the individual maps to the map based on the mean similarity matrix ($M = 0.46$, $SD = 0.10$) were relatively high.

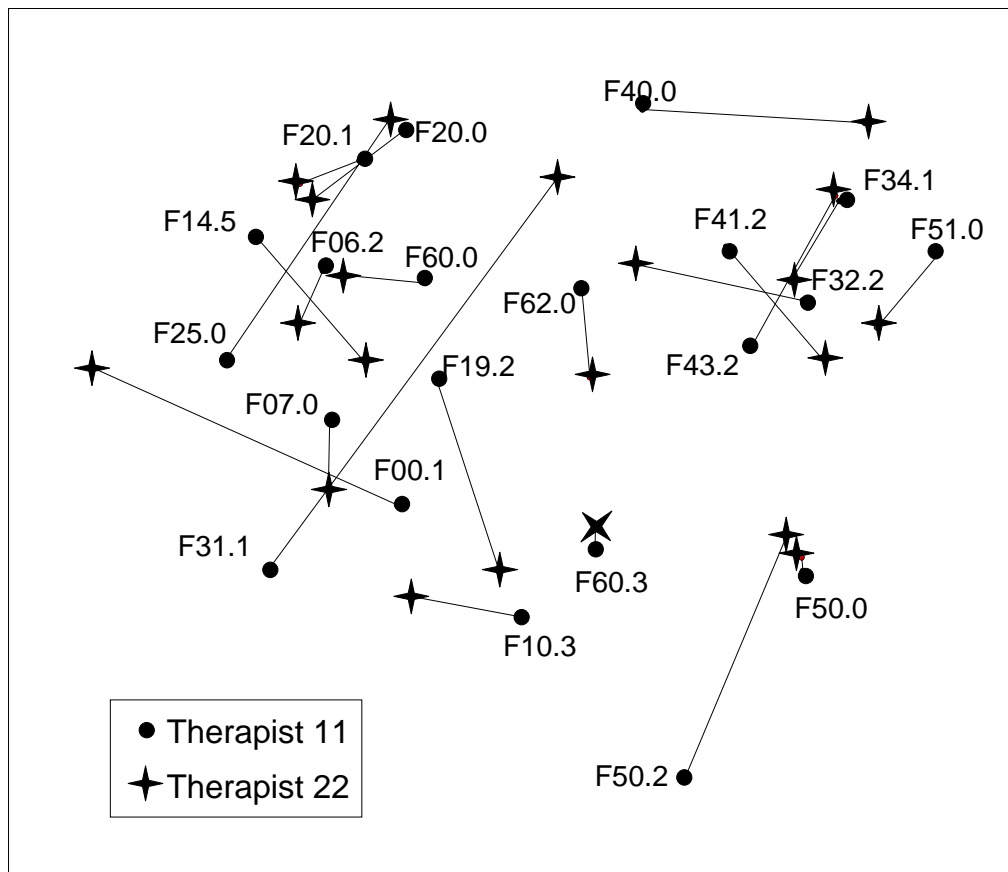


Figure 2. Procrustean Transformation of the cognitive maps of therapist 11 and therapist 22.

The average losses of the comparisons of the individual maps to each other also form the basis for the map presented in Figure 3. Consequently the items in this map represent the relational position of the therapists based on the deviation of their individual cognitive structures from each other (based on only 19/20 therapists, as one outlier was excluded from this analysis). The combination of the additional variables with this map allows it to be examined whether they are associated with the relative positions of the therapists: The abbreviation above each position represents the academic training (psych. = psychology, med. = medical science, both = both degrees), the one below represents the therapeutic approach (CBT = cognitive behavioural, dynam. / analyt. = psychodynamic- / analytic, int. / eclectic = integrative / eclectic and other). In this map, none of the data with regard to professional experience (number of years) and attitude towards the ICD-10 showed significant explicative value (meaningful clusters or significant correlations in the multiple regressions).

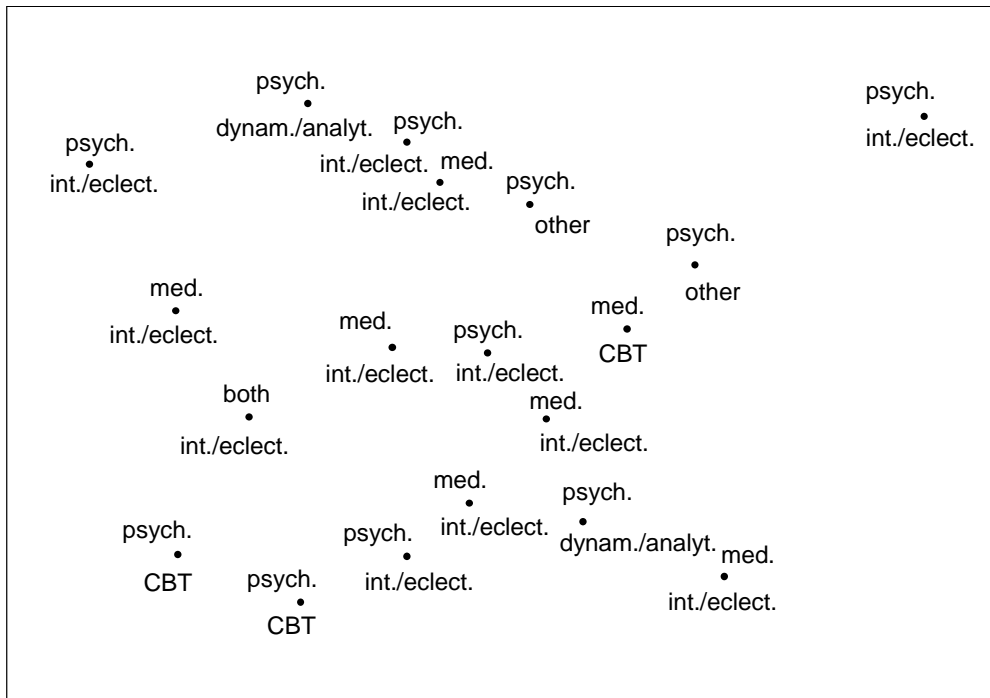


Figure 3. Relational positions of the therapists based on the deviation of their cognitive maps (Academic training: psych. = psychology, med. = medical science, both = both degrees; therapeutic approach: CBT = cognitive behavioral, dynam./analyt. = psychodynamic/analytic, int./eclect. = integrative/eclectic and other).

Consistent with the observation of the therapists that two main aspects in the judgment of similarities of the mental disorders were phenomenology and aetiology, this issue was further investigated. Hence after completing the similarity judgments, they were asked to sort the disorders in groups of maximal similarity according to the criteria mentioned above. Figure 4 shows the map resulting from this parallel sorting task with regard to aetiology (abbr. “a”) and phenomenology (abbr. “p”). Each sorting task resulted in a 0/1 matrix (0 = disorders located in different groups, 1 = disorders located in the same group). The summarized differences of these matrices were used as distance measures to calculate the map (by NMDS). Therefore items which are positioned in spatial proximity of each other represent similar groupings and vice versa. The item labelled “ICD” represents the result which corresponds to the grouping of the disorders according to the ICD-10 F clusters. The delineations subsume all groupings with regard to phenomenology and aetiology respectively.

After having measured the cognitive structures of the mental disorders from the perspective of the therapists by way of similarity judgments and parallel sorting tasks, we verified the clinical plausibility of our analysis by confronting the therapists with the results. This evaluation of the maps and the cluster groups revealed that 8/13 therapists judged the cognitive maps to better represent the structure of the presented mental disorders than the cluster analytic groups. Among the cognitive maps, 10/13 therapists preferred the map based on the mean similarity judgments to the map based on their own individual similarity judgments. The preferred map received a good average rating ($M = 4.6$, $SD = 1.0$) on a Swiss school grade scale (1 = worst, 6 = best) with regard to acceptance by the therapists.

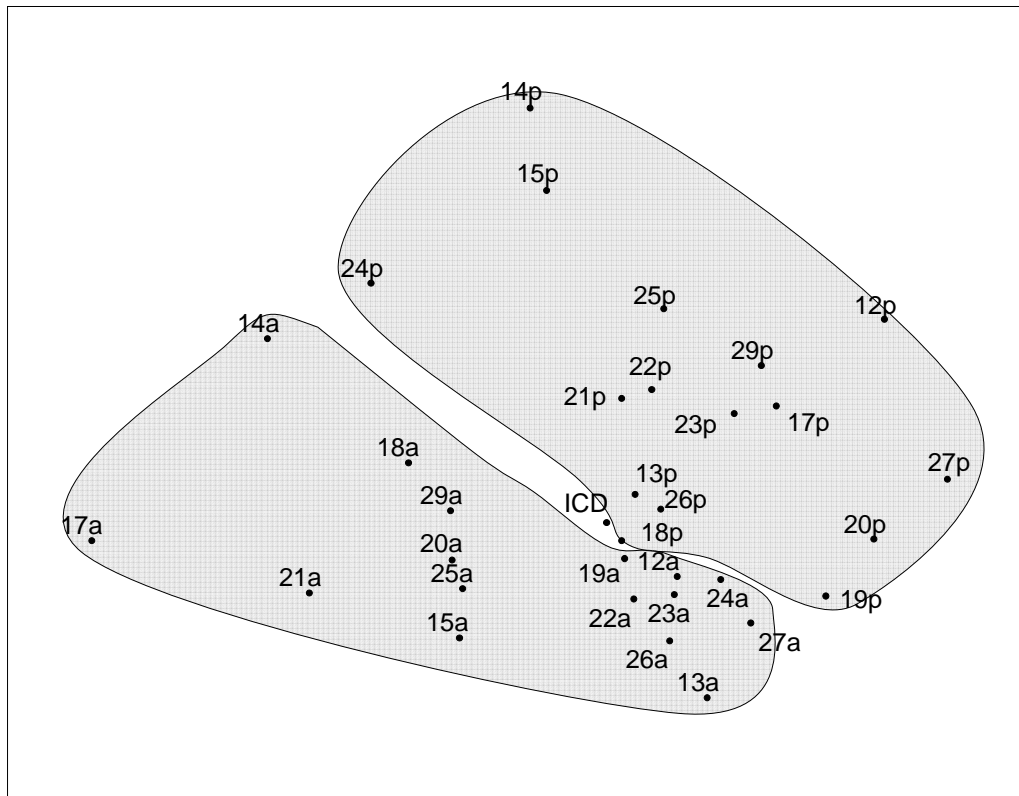


Figure 4. Results of the parallel Sorting task with regard to etiology (a) and phenomenology (p) (ICD = Grouping of the disorders according to the ICD-10 F clusters).

7.4. Discussion

As repeatedly highlighted in earlier studies, the question of whether to look at the classification of mental disorders from a categorical or from a dimensional point of view cannot really be answered exclusively in favour of one of the alternatives (Clark, Watson & Reynolds, 1995; Mundt, 2002; Widiger & Clark, 2000). The present study explored a method which attempts to bear in mind both perspectives from the point of view of therapists, i.e. the individuals who actually have to interpret the classification system. Their structural mental representation of the selected disorders was measured by similarity judgments which were translated into two-dimensional cognitive maps as well as into cluster analytic groups. Considering that the instruction was to judge similarities between the disorders based on their personal clinical experience and not to attempt to reproduce the ICD-10 structure, the spatial representation of the disorders in Figure 1 presents a meaningful configuration.

The plotted clusters depict the ICD-10 F categories, and therefore allow categorical aspects of the structure to be considered. It is noticeable that these clusters were quite exactly reproduced, as they can be clearly delineated from each other, with the exception of three overlapping areas. These areas for their part can be hypothesized to illustrate several of the critical observations regarding the ICD-10 and the DSM-IV classification, which have accumulated in the past years and which are evidently also perceived by the participating therapists. First, the overlap of F3 (mood (affective) disorders) and F4 (neurotic stress-related and somatoform disorders) was already similarly described much earlier by Maser & Cloninger (1990) as the

difficulty of defining a clear-cut boundary between the diagnostic entities of mood and anxiety disorders. Second, the overlap of F6 (disorders of adult personality and behaviour) and F3 is mainly due to the spatial positions of emotionally unstable personality disorder (F60.3) and bipolar affective disorder, current episode manic without psychotic symptoms (F31.1): F60.3 is positioned between the categories of F1 (mental and behavioural disorders due to psychoactive substance use), F3 and F5 (behavioural syndromes associated with physiological disturbances and physical factors – here especially the eating disorders have to be taken into account). It is well known that mental disorders from these three categories are frequently observed as comorbid diagnoses of F60.3 (Gunderson, 2001). F31.1 is located across from the dysthymia (F34.1) and the severe depressive episode without psychotic symptoms (F32.2), and in addition is closely located to schizoaffective disorder, manic type (F25.0) with which it shares most of its symptoms according to the ICD-10. Third, looking at the location of the organic delusional (schizophrenia-like) disorder (F06.2) in the F2 (schizophrenia, schizotypal and delusional disorders) cluster, this position also fits from the point of view of its phenomenological description, as F06.2 is only located in the F0 (organic, including symptomatic, mental disorders) category due to its aetiology.

The structure of the cognitive maps can be interpreted not only with regard to meaningful clusters as demonstrated above, but also in terms of the dimensions: One possibility for this is to collect additional data on external scales with regard to the criteria which are hypothesized to have an explicative value in the explored structure, as was done in a related study by Purkhardt & Stockdale (1993). A considerable number of the selected disorders in our experiment showed a high (cognition) and a fairly high (psychosis) correlation on two of the three dimensions that emerged from the content analysis. In contrast to the dimensions / factors of a factor analysis, these dimensions in the cognitive map calculated by NMDS can be directly attributed to the Euclidian distances between the items / disorders. Thus the spatial position of these disorders in the cognitive map can be partially explained by the underlying dimensions of cognition and psychosis. On the cognition dimension, Dementia in Alzheimer's disease occupies the most extreme position. Therefore this disorder shows the strongest emphasis on this dimension (within the structure of the cognitive map) in the ICD-10 among the selected disorders. The same observation applies to the schizoaffective disorder, which can be found at the most extreme position of the psychosis dimension.

As mentioned in the results section, the reanalysis of the similarity matrix by cluster analysis supports the verification of the clusters found in the cognitive map. The three overlaps which could be found in the map can for the most part also be observed in the results of the cluster analysis (Table 1): Adjustment disorder (F43.2), mixed anxiety and depressive disorder (F41.2), dysthymia (F34.1) and severe depressive episode without psychotic symptoms (F32.2) are subsumed in the same group, and thereby emphasize the overlap of the F3 and the F4 categories. Emotionally unstable personality disorder (F60.3) is placed in the same cluster as the eating disorders, and the organic delusional (schizophrenia-like) disorder (F06.2) is located together with the other schizophrenic disorders. A further interesting congruence of the results of the cluster analysis and the cognitive map can be observed in terms of the spatial extension of the clusters: For instance, the F6 (disorders of adult personality and behaviour)

cluster, which shows a large circumference in the map, can be found in three different groups in the cluster analysis. The F1 (mental and behavioural disorders due to psychoactive substance use) cluster on the other hand, which subsumes its disorders in a small area in the cognitive map, can also be found in only one group in the cluster analysis. These findings could be interpreted as an indicator that the selected disorders belonging to a large cluster can be less clearly distinguished from other clusters than disorders belonging to a small cluster. Additionally, according to Blashfield (1990), large clusters are an indicator for a dimensional rather than a categorical structure and vice versa. The latter observation with regard to the F6 cluster is also supported by suggestions in the literature for a dimensional structure of personality disorders (Mundt, 2002; Oldham & Skodol, 2000; Skodol, Gunderson, Pfohl, Widiger, Livesley & Siever, 2002).

After having discussed the cognitive map based on the mean similarity matrix it is also interesting to look at the individual maps. The rather high average losses (i.e. the extent of the deviation) of the individual maps compared to each other as well as compared to the map based on the mean similarity judgments, suggest that the therapists' cognitive structures were highly individual. Cognitive maps of experts in other domains of expertise (e.g. ornithology) showed lower average losses (approx. 0.4) when compared to each other (Laege, 2001). Figure 2 illustrates the deviating relative positions of the corresponding disorders in the cognitive maps of the two therapists 11 and 22 chosen as an example. It is evident that some disorders show only a small distance between each other (e.g. the personality disorders) and are therefore conceptualized similarly by the therapists. Other disorders (e.g. bipolar affective disorder, current episode manic without psychotic symptoms (F31.1) and Dementia in Alzheimer's disease with early onset (F00.1)) show a larger discrepancy, which can be interpreted in terms of a larger divergence of these disorders in the cognitive structure of the corresponding therapist compared with the other therapist. Despite the considerable divergence of the subjects in our sample with regard to academic training, clinical experience and therapeutic approach (as described in the methods section), these variables did not show any explicative value in the map representing the relational position of the therapists (i.e. their cognitive maps). In this regard it is noteworthy that neither the often presumed gap between psychologists and psychiatrists, nor a grouping of the therapists in terms of their therapeutic approach could be confirmed in this study (Figure 3).

Another reason for the considerable discrepancies of the individual structures might be found in the variability of the similarity judgments. As they were based on personal clinical experience, they may involve multiple criteria, e.g. aetiology, phenomenology, comorbidity etc. This argument was explored in a successive experiment by constricting the similarity judgments to a specific criterion (i.e. phenomenology). This resulted in a smaller variance and a lower level of the average losses between the maps (Streule, Egli, Schlatter & Laege, 2006). The ability of the therapists to constrict their judgment to a single criterion (i.e. aetiology and phenomenology) was also examined in the parallel sorting task. The map presented in Figure 4 illustrates that the therapists were able to differentiate the two tasks well, as the corresponding results of the phenomenology task (abbr. "p") can clearly be delineated from the results of the aetiology task (abbr. "a"). Interestingly, the result based on the ICD-10 grouping is not

located in the phenomenology cluster as could be expected (Stieglitz, Freyberger & Mombour, 2002) but rather between the phenomenology and the aetiology clusters. Thus this map illustrates the conflict of classification between these two criteria, which is evidently also perceived by the therapists who participated in our experiment.

In conclusion, the cognitive maps based on the similarity judgments offer an opportunity to empirically explore and visualize the structure of the mental disorders directly from the point of view of the therapists, who are the individuals who ultimately have to interpret and use a classification system such as the ICD-10. As we have demonstrated, a cognitive map allows the integration of a categorical as well as a dimensional perspective. The evaluation by the therapists showed that the majority favoured this presentation over a pure categorical presentation such as cluster analysis, and deemed it to be of good quality. The map based on the mean similarity matrix was preferred over the maps based on the individual matrices, and the individual maps showed a smaller divergence when compared to the mean map than compared to each other. These observations indicate that a common structure of the disorders of the participating therapists is a plausible assumption and this common configuration of the disorders as a cognitive map can therefore serve as an expert model. Such an expert model can be used for instance to structure and monitor learning progress in the field of psychopathology, where it is already being tested and included in an extensive e-learning project. Additionally, it could open up opportunities to explore local differences in perceptions of therapists regarding a common diagnostic system (thereby serving as a new approach to determine interrater reliabilities already on the level of the perception of the classification system) as well as to refine the discrepancies and convergences between related diagnostic systems such as the ICD and the DSM. Furthermore profiles of patients (e.g. descriptions of case examples) could be fitted into a cognitive map of mental disorders and help to clarify issues of comorbidity and multiple diagnoses. Combined with the arguments mentioned above this method might help to build a bridge across the gap between theoretical concepts and clinical practice in the field of psychopathology.

7.5. References

- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders, ed. 4 (DSMIV)*. Washington: American Psychiatric Association.
- Arbeitsgemeinschaft fuer Methodik und Dokumentation in der Psychiatrie (2000). *Das AMDP-System Manual zur Dokumentation psychiatrischer Befunde* (7. Aufl.). Goettingen: Hogrefe.
- Blashfield, R. K. (1990). Comorbidity and classification. In J. D. Maser & C. R. Cloninger (Eds.), *Comorbidity of mood and anxiety disorders* (pp. 61-82). Washington: American Psychiatric Press.
- Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling – Theory and applications*. New York: Springer.
- Bortz, J. (1999). *Statistik fuer Sozialwissenschaftler* (5. Aufl.). Berlin: Springer.
- Clark, L. A., Watson, D. & Reynolds, S. (1995). Diagnosis and Classification of Psychopathology: Challenges to the Current System and Future Directions. *Annual Review of Psychology*, 46, 121-153.
- Coxon, A. P. M. (1982). *The users guide to multidimensional scaling*. London: Heinemann.
- Daub, S (2001). *Similarity Simulation - wie man den Code des Globalurteils knackt*. Dissertation, Philosophische Fakultät, Universität Zuerich. Zuerich.

- Dittmann, V. (1996). Grundprinzipien der psychiatrischen Diagnostik und Klassifikation. In B. Neundoerfer (Hrsg.), *Atlas der Nervenheilkunde Neurologie und Psychiatrie in Bild und Wort* (S. 282-285). Karlsruhe: Braun Fachverlage.
- Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie*. Muenchen: Reinhardt.
- Gunderson, J. G. (2001). *Borderline personality disorder – A clinical guide*. Washington: American Psychiatric Press.
- Hurley, J.R. & Cattell, R. B. (1962). The Procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, 7, 258-262.
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32, 241-254.
- Laege, D. (2001). *Aehnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift, Philosophische Fakultät, Universität Zuerich. Zuerich.
- Laege, D., Daub, S., Bosia, L., Jäger, C. & Ryf, S. (2005). *Die Behandlung ausreißer-behafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag*. AKZ-Forschungsbericht Nr. 21. Zürich: Angewandte Kognitionspsychologie.
- Lange, W., Munk-Jorgensen, P., Bertelsen, A., Schurmann, A., Michels, R., Malchow, C. P. & Dilling, H. (2002). Comparison of psychiatric ICD-10 diagnoses in Denmark and Germany. *Psychopathology*, 35, 36-47.
- Maser, J. D. & Cloninger, C. R. (Eds.) (1990). *Comorbidity of mood and anxiety disorders*. Washington: American Psychiatric Press.
- Merten, K. (1995). *Inhaltsanalyse: Einfuehrung in Theorie, Methode und Praxis* (2. Aufl). Opladen: Westdeutscher Verlag.
- Microsoft Corporation (2002). *Microsoft Excel v 2002*. Microsoft Corporation.
- Mundt, C. (2002). Psychological perspectives for the development of future diagnostic systems. *Psychopathology*, 35, 145-151.
- Mussigbrodt, H., Michels, R., Malchow, C.P., Dilling, H., Munk-Jorgensen, P. & Bertelsen, A. (2000). Use of the ICD-10 classification in psychiatry: An international survey. *Psychopathology*, 33, 94-99.
- Neuenschwander, M., Meyer, P. C. & Hell, D. (1999). *Tabellenband 1998: stationaere Behandlungen in psychiatrischen Kliniken, Psychotherapiestationen und Institutionen fuer Suchtkranke des Kantons Zuerich*. Psychiatrische Universitätsklinik Zuerich. Zuerich.
- Oldham, J. M. & Skodol, A. E. (2000). Charting the future of axis II. *Journal of Personality Disorders*, 14, 17-29.
- Purkhardt, S. C. & Stockdale, E. (1993). Multidimensional scaling as a technique for the exploration and description of a social representation. In G. M. Breakwell & D. V. Canter (Eds.), *Empirical Approaches to Social Representation* (pp. 272-297). Oxford: Clarendon Press.
- Ryf, S., Daub, S. & Laege, D. (2003). *INTUS - Computergestuetzte Erfassung von relationalen Urteilen*, v 2.4.4. Allgemeine Psychologie, Psychologisches Institut, Universität Zuerich. Zuerich.
- Schnyder, F., Daub, S., Jaeger, C., Ryf, S., Christen, S. & Laege, D. (2003). *ASSO - Statistische Auswertung von Assoziationsdaten*, v 3.2.3. Allgemeine Psychologie, Psychologisches Institut, Universität Zuerich. Zuerich.
- Skodol, A. E., Gunderson, J. G., Pfohl, B., Widiger, T. A., Livesley, W. J. & Siever, L. J. (2002). The borderline diagnosis I: Psychopathology, comorbidity and personality structure. *Biological Psychiatry*, 51, 936-950.
- Smith, E. E., Shoben, E. J. & Rips, L. J. (1974). Structure and process in semantic memory. A featural model for semantic decisions. *Psychological Review*, 81, 214-241.
- SPSS Inc. (2001). *SPSS for Windows v 11.0.1.*. SPSS Inc..
- Stieglitz, R. D., Freyberger, H. J. & Mombour, W. (2002). Klassifikation und diagnostischer Prozess. In H. J. Freyberger, W. Schneider & R. D. Stieglitz (Hrsg.), *Kompendium Psychiatrie Psychotherapie Psychosomatische Medizin* (11. vollst. ern. u. erw. Aufl.) (S. 17-31). Basel: Karger.
- Streule, R., Egli, S., Schlatter, K. & Laege, D. (2006). *The structure-based expert model of the mental disorders – A validation study*. Manuscript in preparation.

- Widiger, T. A. & Clark, L. A. (2000). Toward DSM-V and the classification of psychopathology. *Psychological Bulletin*, 126, 946-963.
- Wiedenbeck, M. & Zuell, C. (2001). Klassifikation mit Clusteranalyse: Grundlegende Techniken hierarchischer und K-means-Verfahren. *ZUMA How-to-Reihe*, 10.
- World Health Organization (1992). *ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

8.

**Das Zürcher
Wissensstrukturmodell für
Psychopathologie (ZWP)
—
eine Pilotstudie**

*Zusammenfassung*¹³

Theoretischer Hintergrund: Es besteht ein Bedarf an objektiven Kriterien zur Evaluation von psychotherapeutischen Aus-, Weiter- und Fortbildungsprogrammen.

Fragestellung: Überprüfung der Verfahrenslogik „Das Zürcher Wissensstrukturmodell für Psychopathologie (ZWP)“ als Qualitätssicherungsinstrument.

Methode: Darstellung der strukturellen Aspekte des individuellen deklarativen therapeutischen Sachwissens (psychopathologische Störungsbilder) in Wissenskarten über die Nonmetrische Multidimensionale Skalierung und Vergleich mit einem Normmodell zur Abschätzung der momentanen Wissensqualität mit 26 Kandidaten aus den vier Jahrgängen der postgradualen psychotherapeutischen Weiterbildung (Verhaltensmedizin und Kognitive Verhaltenstherapie) der Universität Zürich.

Ergebnisse: Mit zunehmender Expertise unterscheiden sich die Wissenskarten der Auszubildenden immer weniger von einem Expertenmodell, und auch die Unterschiede zwischen den Karten der Auszubildenden nehmen ab.

Schlussfolgerung: Die Resultate sprechen für den Einsatz von ZWP als komplementäres Evaluationsinstrument in der psychotherapeutischen Aus-, Weiter- und Fortbildung.

Schlüsselwörter: Nonmetrische Multidimensionale Skalierung, Prokrustes-Transformation, Qualitätssicherung, Bildungsprogramme, Evaluationsinstrument, Wissensmessung, Psychotherapie-Training, Psychopathologie

8.1. Einleitung

„Die Qualität der therapeutischen Versorgung kann nur so gut sein wie die Ausbildung der Therapeuten.“ lautet der einleitende Satz des Beitrags von Reinecker & Schindler (2000, S. 655). Diese Feststellung führt zur Forderung nach einer qualitativ hochwertigen, sich ständig weiterentwickelnden und an neue Erkenntnisse aus der psychotherapeutischen Forschung anpassenden Aus-, Weiter- und Fortbildung. Gleichzeitig ist damit aber auch die Frage der Evaluation der eingesetzten Lehr- und Lerninhalte sowie der didaktischen Methoden angesprochen. Trotz der Forderung nach effektiven Psychotherapie-Bildungsprogrammen fehlt es vielfach an überzeugenden Evaluationsinventaren.

Kenntnisse (Wissen um theoretische Grundlagen), Fertigkeiten (konkreter Umgang mit Patienten) und Erfahrung (Aspekte der Interaktion eigener persönlichen Merkmale mit der therapeutischen Situation) bilden die drei Grundpfeiler der Kompetenz qualifizierter Psychotherapeuten (Buchkremer & Schneider, 1995; Reinecker & Schindler, 2000). Diese zu erreichen ist

¹³ An dieser Stelle möchten wir uns bei Frau Prof. Dr. Ulrike Ehlert und Frau Dr. Nora Kaiser für die Möglichkeit zur Durchführung der Studie im Rahmen ihres Weiterbildungskurses sowie für die anregenden Gespräche zum Einsatz des Verfahrens und zur Einordnung der Resultate bedanken.

das explizit zu fordernde Ergebnis eines Bildungsprogramms (Willutzki & Botermans, 1997). Dabei wird diese Qualifikations-Charakteristik üblicherweise als gestufter Prozess betrachtet: Viele Aus-, Weiter- und Fortbildungen in Psychotherapie vermitteln zu Beginn des Lernprozesses grundlegendes Wissen über Störungsbilder und Handlungsmodelle. Erst dann folgen praktische Anwendungen, welche Fertigkeiten und die notwendige (Selbst)Erfahrung vermitteln. In diesem gestuften Lernprozess soll die Umsetzung abstrakter Wissensinhalte in automatisierte und adäquate Handlungsabläufe das Ziel der psychotherapeutischen Bildung sein (Eckert & Westhof, 2000, Freedheim & Overholser, 1998). Grundlage und zwingend notwendige Voraussetzung für ein an die jeweilige Situation angepasstes therapeutisches Verhalten ist damit das Vorhandensein deklarativer, kognitiver Wissensstrukturen. Erst die differenzierte Verarbeitung, das Erkennen von Gemeinsamkeiten und Unterschieden, die Integration und Elaboration einzelner Wissensinhalte in ein Gesamtbild in der Ausbildung ermöglichen den anschließenden Wissenstransfer in die praktische Situation hinein. Das dürfte der Grund sein, warum auch in der praktisch orientierten Therapieausbildung die Vermittlung vernetzten Wissens eine zentrale Funktion wahrnimmt.

Die Ergebnisqualität eines an solchen Zielen orientierten psychotherapeutischen Curriculums sicherzustellen, ist freilich nicht trivial. Schwierigkeiten ergeben sich bereits bei der Bestimmung objektiver Standards zur Beurteilung von Therapieausbildungen: Auch wenn intuitiv ein Verständnis darüber vorhanden ist, was einen guten Therapeuten ausmacht (Reinecker & Schindler (2000) sowie Freedheim & Overholser (1998) listen eine Reihe von Aspekten auf), bleibt die Operationalisierung dieser Faktoren ein kritisches Unterfangen. Auffällig oft werden Verfahren der umfassenden Selbsteinschätzung zur Evaluation von Bildungsprogrammen verwendet, wobei sich die Datengrundlage meist auf ein deskriptives Niveau beschränkt. Selbsteinschätzungen behelfen sich beispielsweise der Erfassung von Merkmalen wie der allgemeinen Zufriedenheit mit der Ausbildung (Foulkes, 2003), des subjektiv beurteilten Lernfortschritts (Schulz, 2000; Schulz et al. 1998), der allgemeinen Erweiterung bzw. Steigerung des Selbstvertrauens (Selbsteinschätzung der therapeutischen Kompetenz: Beitman & Yue, 1999; Elizur et al., 1994; Jagsch et al., 2002; Milne et al., 1999; Schulz, 2000; Schulz et al., 1998; Willutzki & Botermans, 1997) oder des auf die Aus- bzw. Weiterbildung zurückzuführenden beruflichen Aufstiegs (Schulz, 2000; Schulz et al., 1998).

Diese – zumeist über Fragebögen oder Interviews erhobenen – Selbsteinschätzungen bilden für den Ausbildungspfeiler der Erfahrung und, mit Abstrichen, auch für den der Fertigkeiten sicherlich die geeignete Datengrundlage für eine Evaluation des Erreichten. Sofern es aber um die Messung des vernetzten Wissens geht, welches ja als Grundlage für die weiteren, praktischen Ausbildungsschritte gesehen wird, sollten objektivere Evaluationskriterien den Bildungsprozess begleiten. Zu schnell wird das eigene deklarative Fachwissen als adäquat vernetzt eingeschätzt, denn der aktuelle Wissensstand definiert ja das, was als Wirklichkeit erlebt (und damit als Vergleichskriterium herangezogen) wird. Aus diesem Grund ist es positiv zu sehen, dass die meisten Bildungsinstitutionen für den Bereich des deklarativen Wissens Prüfungen verlangen.

Leider fragen derartige Prüfungen – sofern sie standardisiert sind – in aller Regel nur einzelne Details ab, und das notwendigerweise auch nur stichprobenartig. Die Vernetzung des Wissens

(d.h. das Sehen von Zusammenhängen und Unterschieden) und die Wissensstruktur als Ganze lassen sich auf diese Weise aber nur unzureichend überprüfen. Aus diesem Grund bietet es sich an, ergänzend zum traditionellen Abfragen von Fakten auch relationales Wissen zu erheben, und zwar auf eine genauso standardisierte und damit vergleichbare Weise wie es die üblichen Prüfungsformate ermöglichen.

Die im Folgenden vorgestellte Studie wird dazu einen methodischen Vorschlag unterbreiten. Sie stützt sich exemplarisch auf die Messung des relationalen Wissens über psychopathologische Störungsbilder, welches, wie Reinecker & Schindler (2000) anmerken, einen wichtigen Pfeiler therapeutischer Kompetenz bildet. Das störungsspezifische Wissen mit den Dimensionen Phänomenologie, Ätiologie, Prävalenz usw. umfasst eine Vielzahl an einzelnen Aspekten, die vom Lernenden im Laufe der Ausbildung zueinander in Beziehung zu setzen und zu einem gesamthaften relationalen Wissensbild zu integrieren sind. Deswegen erscheint es vertretbar, das vorgeschlagene Evaluationsverfahren an diesem Inhaltsbereich des deklarativen therapeutischen Wissens zu entwickeln und erstmalig zu überprüfen.

Das Verfahren folgt der strukturellen Idee von Mapping-Techniken (vgl. Mandl & Fischer, 2000), überlässt die Anforderung des gesamthaften Strukturierens aber nicht dem Befragten (was nur zu oft zu Überforderungen führt), sondern zerlegt die Gesamtaufgabe des Wiedergebens relationalen Wissens in kleinstmögliche Einheiten. Dazu erhebt es über numerische relationale Urteile die subjektiv eingeschätzte Ähnlichkeit zwischen je zwei psychopathologischen Störungsbildern, wobei das Verfahren bewusst die Kriterien für die Ähnlichkeitseindrücke dem Wissensstand der evaluierten Person überlässt und nicht durch das Frageformat vorgibt. Im Laufe einer längerfristigen Ausbildung können sich diese Kriterien nicht nur erweitern, sondern auch fundamental verändern (in der Tradition Piagets (1976) würde das dann als Akkomodationsleistung bezeichnet).

Die durch numerische Ähnlichkeitsurteile ermittelten Proximitäten zwischen Paaren von Objekten lassen sich statistisch durch Nonmetrische Multidimensionale Skalierung (NMDS) als Distanzwerte verrechnen und so zur Grundlage von Wissenskarten machen (die Logik dieses Verfahrens ist z.B. bei Borg & Groenen, 1997 beschrieben; vgl. auch Läge, Daub, Bosia, Jäger & Ryf, 2005). Auf diese Weise gewinnt man die relationale Information über das aktuelle Sachwissen auf personenspezifischer Ebene. Solche individuellen Wissenskarten können durch Prokrustes-Transformation mit einem Expertenmodell verglichen werden (Gower & Dijksterhuis, 2004; Läge, 2001). Bereits ohne großen mathematischen Aufwand springt ins Auge, welche Objekte der Lerner gut kennt (d.h. korrekt verortet) und welche er falsch platziert (Abbildung 60). Die Abweichungen zwischen der Lernerkarte und der Expertenkarte werden auf der Objektebene über die Distanz zwischen den beiden korrespondierenden Punkten (= Object Loss oder kurz ObjLoss) quantifiziert. Die gemittelte Summe aller Einzelabweichungen ergibt das Gesamtmaß des Average Loss (kurz als AvgLoss bezeichnet). Der AvgLoss dient also der Beurteilung des gesamten Wissensstandes einer einzelnen Person, verglichen mit einem Norm- / Expertenmodell.

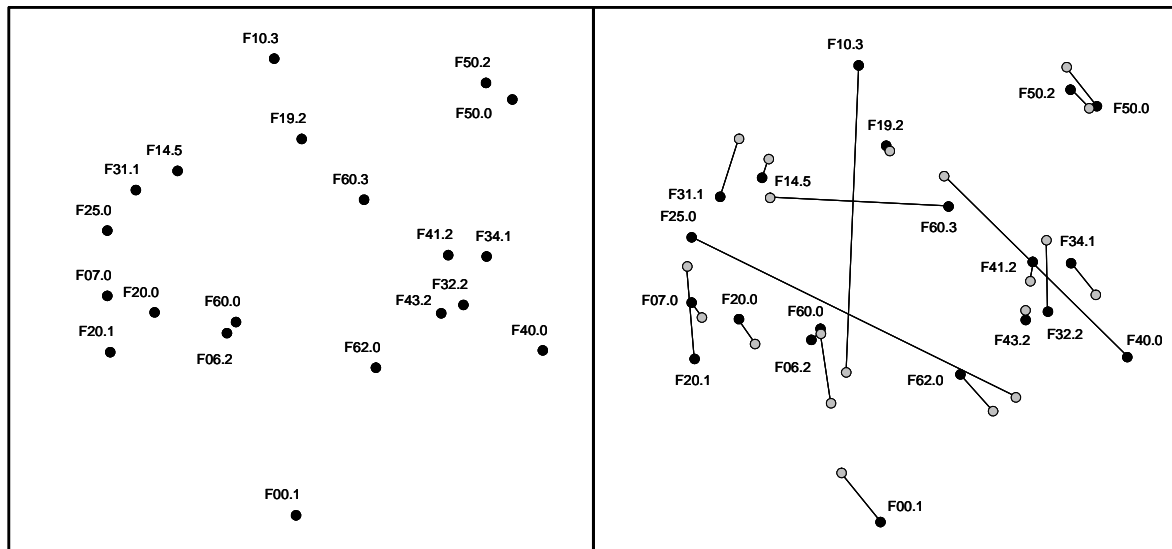


Abbildung 60. Auf Ähnlichkeitsurteilen beruhende NMDS-Lösung (links) und Ergebnis einer Prokrustes-Transformation (rechts). Die linke Karte steht für das in dieser Studie verwendete Expertenmodell. Abgebildet sind 20 Störungsbilder, codiert nach ICD-10 (s. Tabelle 4). Die grauen Punkte in der rechten Abbildung stehen für die Objektpositionen einer Lerner-Karte. Die korrespondierenden Punkte zwischen der Normkarte (links) und der Lernerkarte sind mit einer Verbindungslinie gekennzeichnet (= ObjLoss). Die Gesamtabweichung beträgt in diesem Beispiel $\text{AvgLoss} = 0.37$. Eine strukturelle Ähnlichkeit der beiden Karten ist erkennbar, jedoch weisen einige Objektpositionen des Lerners große Abweichungen auf (z.B. F25.0, F10.3, F40.0).

Der Erstellung des Vergleichskriteriums (Normmodell) muss großes Gewicht beigemessen werden. Im Gegensatz zu anderen Inhaltsgebieten, in denen Ähnlichkeitsrelationen zwischen Objekten über die mathematische Berechnung bestimmt werden können, werden im hier vorgestellten Bereich der psychopathologischen Störungsbilder die Sichtweisen mehrerer Fachpersonen empirisch erfragt und zu einem gemeinsamen Modell verrechnet. Damit verschwinden individuelle Gewichtung Unterschiede in den Beurteilungsdimensionen und das „Gemeinsame“ in den psychopathologischen Weltbildern dieser Experten kristallisiert sich heraus. Relevanz und Validität einer solchen Mittelwertsbildung lässt sich aus den Resultaten kontrollierter Lernexperimente herauslesen, in denen gezeigt werden konnte, dass über das eben beschriebene Verfahren Faktenwissen und Wissensveränderungen über einen kontrollierten und für alle Probanden gleichsam strukturierten Lernprozess im Sinne einer Konvergenz individueller Wissenskarten an ein Expertenmodell gemessen werden können (vgl. Kapitel 2-4 in dieser Arbeit). Das in der hier präsentierten Studie verwendete Normmodell (Abbildung 60, links) stammt aus einer Expertenbefragung von Egli, Schlatter, Streule & Läge (2006). 20 praktizierende Therapeuten (8 w, 12 m) aus der Umgebung von Zürich hatten dazu die Ähnlichkeit dieser Störungsbilder paarweise eingeschätzt. Die gemittelten Ähnlichkeitsdaten wurden mit dem robusten NMDS-Algorithmus RobuScal (Läge, Daub, Bosia, Jäger & Ryf, 2005) skaliert und von den Experten qualitativ evaluiert. Die Auswahl der Störungsbilder erfolgte zum einen aufgrund der Repräsentativität (aus allen F-Kategorien des ICD-10 ohne F7, F8 und F9) und Variabilität und zum anderen aufgrund der klinischen Relevanz (Prävalenz).

Im Verlauf einer psychotherapeutischen Ausbildung sollten sich die individuellen strukturellen Sichtweisen nach und nach an das Bild annähern, welches als gemeinsame Essenz der

Expertensicht gesehen werden kann. Diese Hypothese folgt zumindest aus dem typischen konvergenten Lernstoff einer Ausbildung, und das Erreichen einer Annäherung setzt zugleich voraus, dass einzelne gelernte Fakten nicht nur abgespeichert, sondern zu einem relationalen Wissen verknüpft (also integriert und elaboriert) sein sollte. Konkret auf den gewählten Objektbereich bezogen bedeutet dies, dass die adäquate Verarbeitung der Merkmalsinformation es einer Person erlaubt, Unterschiede und Gemeinsamkeiten zwischen Störungsbildern zu entdecken. Dieses wiederum sollte in letzter Konsequenz zu differenzierten und qualitativ stimmigen Ähnlichkeitsurteilen führen. Mit zunehmendem Expertisegrad (in der folgenden Pilotstudie operationalisiert über das Weiterbildungsjahr) sollten sich somit die Wissensunterschiede (operationalisiert über die AvgLoss der individuellen Lernerkarten vom Normmodell) verkleinern. Des Weiteren ist zu erwarten, dass sich durch die Auseinandersetzung aller Auszubildenden mit dem gleich strukturierten Lehrgang eine Homogenisierung der Wissenskarten mit zunehmendem Expertisegrad ergibt. Die individuellen Karten sollten sich also mit fortschreitender Ausbildungszeit auch aufeinander zu bewegen.

Diese Verfahrenslogik des Vergleichs zwischen Lernerkarten und Expertenkarten wird in der vorgestellten Pilotstudie nicht nur erstmalig für das Inhaltsgebiet der Psychopathologie überprüft, sondern es handelt sich überhaupt um den ersten probenhaften Einsatz dieses Evaluationsinstrumentes im „real existierenden“ Unterricht (d.h. außerhalb von kontrollierten Laborexperimenten). Sollten sich die aufgestellten Hypothesen einer systematischen Konvergenz der gemessenen Wissenskarten bestätigen lassen, so läge damit ein Indiz vor, dass der Ausbau dieses Verfahren für einen breiteren Praxiseinsatz lohnenswert sein könnte.

8.2. Methode

Stichprobe. Eine erste Pilotstudie wurde in Kooperation mit dem Lehrstuhl Klinische Psychologie und Psychotherapie der Universität Zürich im Rahmen der postgradualen psychotherapeutischen Weiterbildung mit den Schwerpunkten Verhaltensmedizin und Kognitiver Verhaltenstherapie¹⁴ (Ehlert & Kaiser, 2001; Kaiser & Ehlert, 2003) im Wintersemester 03/04 durchgeführt. Der Personenkreis besteht aus Psychologen und Ärzten. Die 4-jährige, berufsbegleitende Weiterbildung für Psychologen ist von der Föderation Schweizer Psychologinnen und Psychologen (FSP) anerkannt.

Von den 37 Teilnehmern in den vier Weiterbildungsjahren konnten insgesamt 26 Personen (24 w, 2 m) für die freiwillige Teilnahme an der Untersuchung gewonnen werden. Aus dem ersten Weiterbildungsjahr stammen 14 (von 15) Personen (13 w, 1 m), aus dem zweiten 8 (von 8) Personen (8 w), aus dem dritten 2 (von 9) Personen (1 w, 1 m) und aus dem vierten 2 (von 5) Personen (2 w). Die einzelnen Stichproben werden in der Folge mit WB1 (= 1. Weiterbildungsjahr) bis WB4 (= 4. Weiterbildungsjahr) gekennzeichnet.

¹⁴ Als Lehrinhalt werden im 1. und 2. Weiterbildungsjahr Grundlagen, theoretisches Basiswissen und störungsspezifisches Wissen vermittelt, durchgängig begleitet von Selbsterfahrung und klinischer Praxis. Im weiteren Verlauf (2. - 4. Weiterbildungsjahr) kommen eigene therapeutische Erfahrungen, Gruppensupervision und Fallseminare hinzu. Mehrere Experten von verschiedenen Institutionen dozieren zu ihren jeweiligen Fachgebieten.

Erhebungsdesign. Das Untersuchungsdesign ist als Quasi-Längsschnitt über die vier Weiterbildungsjahrgänge hin konzipiert. Die computerbasierte Datenerhebung fand in der zweiten Woche des Wintersemesters 03/04 statt. Die ersten beiden probandenstarken Weiterbildungsjahrgänge (WB1, WB2) führten den Versuch im Plenum in Laborräumen durch; die Datenerhebung mit den vier Versuchspersonen des dritten und vierten Jahrgangs (WB3, WB4) erfolgte individuell in Laborräumen. Die Probanden führten die Ähnlichkeitseinschätzung der 20 ausgewählten Störungsbilder (Tabelle 4) nach dem Beurteilungskriterium Phänomenologie durch (der höheren Kontrollierbarkeit wegen wurde auf andere Beurteilungsdimensionen wie Ätiologie usw. zunächst einmal verzichtet). Alle 190 Paare von Störungsbildern wurden in randomisierter Reihenfolge am Computer dargeboten und waren jeweils auf einer neunstufigen Ähnlichkeitsskala (1 = minimale Ähnlichkeit; 9 = maximale Ähnlichkeit) zu beurteilen. Die Befragten benötigten für diese Aufgabe zwischen 15 und 25 Minuten.

Einschränkende Rahmenbedingungen. Der Rückgriff auf ein bereits vorliegendes Expertenmodell bedeutet, dass nicht alle vorkommenden Störungsbilder in gleicher Weise für die postgraduale psychotherapeutische Weiterbildung mit den Schwerpunkten Verhaltensmedizin und Kognitiver Verhaltenstherapie der Universität Zürich relevant sind. Die in der Studie von Egli et al. (2006) angefragten Experten konnten zudem die Beurteilungsdimensionen selbst wählen (im Gegensatz zur ausdrücklichen Aufforderung an die Teilnehmer der postgradualen Weiterbildung, sich bei den Beurteilungen an das Kriterium Phänomenologie zu halten). Nachbefragungen der Experten hatten zwar eine starke Fokussierung auf dieses Kriterium zu Tage gebracht, ein Einfluss anderer Beurteilungsdimensionen auf das Ähnlichkeitsurteil ist jedoch nicht auszuschließen. Diese partielle Inkompatibilität von Expertenmodell und Lernzielen der Weiterbildung führt dazu, dass die hier vorgelegte Pilotstudie zwar das zentrale Testkriterium überprüfen kann (nämlich ob die Weiterbildung zu einer systematischen Annäherung der individuellen Kognitiven Karten an ein allgemeines Expertenmodell führt), dass sie jedoch die tatsächliche Leistungsfähigkeit dieses Verfahrens als begleitende Längsschnittdiagnostik systematisch unterschätzen dürfte.

Tabelle 4. Diagnostizierte Wissensobjekte: 20 psychopathologische Störungsbilder aus der ICD-10-Kategorie F. Linke Spalte: F-Codes; Rechte Spalte: ausführliche Bezeichnung.

F00.1	Demenz bei Alzheimer-Krankheit mit spätem Beginn
F06.2	organische wahnhafte (schizophreniforme) Störung
F07.0	organische Persönlichkeitsstörung
F10.3	Störung durch Alkohol, Entzugssyndrom
F14.5	Störung durch Kokain, psychotische Störung
F19.2	Störung durch multiplen Substanzgebrauch und Konsum sonst. psychotr. Subst., Abhängigkeitssyndrom
F20.0	paranoide Schizophrenie
F20.1	hebephrene Schizophrenie
F25.0	schizomanische Störung
F31.1	bipolare affektive Störung, gegenwärtig manische Episode, ohne psychotische Symptome
F32.2	schwere depressive Episode ohne psychotische Symptome
F34.1	Dysthymia
F40.0	Agoraphobie
F41.2	Angst und depressive Störung, gemischt
F43.2	Anpassungsstörung
F50.0	Anorexia nervosa
F50.2	Bulimia nervosa
F60.0	paranoide Persönlichkeitsstörung
F60.3	emotional instabile Persönlichkeitsstörung
F62.0	andauernde Persönlichkeitsänderung nach Extrembelastung

8.3. Resultate

Interpretierbarkeit der Wissenskarten. Die individuellen Ähnlichkeitsmatrizen wurden mit dem NMDS-Algorithmus RobuScal (Läge, Daub, Bosia, Jäger & Ryf, 2005) zweidimensional skaliert. Zur Überprüfung der statistischen Qualität der resultierenden NMDS-Karten wird der NMDS-Stress herangezogen (vgl. Borg & Groenen, 1997). Dieser Wert steht für die Abbildungsgüte der einer Karte zugrunde liegenden Ähnlichkeitseinschätzungen im Vergleich zu den relativen Verhältnissen der Objektpositionen in der Karte. Die Werte der individuellen NMDS-Lösungen bewegen sich durchgängig unter dem von Gigerenzer (1981, S. 353) für eine zweidimensionale, aus 20 Objekten bestehende NMDS-Lösung als kritisch erachteten Niveau von 0.28 (WB1: $M = 0.16$, $SD = 0.02$; WB2: $M = 0.15$, $SD = 0.02$; WB3: $M = 0.16$, $SD = 0.01$; WB4: $M = 0.14$, $SD = 0.03$). Die individuellen Wissenskarten bauen sich somit aus systematisch zueinander passenden Ähnlichkeitsurteilen auf. Das bedeutet, bereits im ersten Jahr geben die Kursteilnehmer in sich hinlänglich konsistente Urteile ab (was natürlich nichts über die inhaltliche Qualität aussagt). Damit ist eine wichtige Grundvoraussetzung gegeben, die relationalen Ähnlichkeitsurteile überhaupt als Messinstrument für das strukturelle Wissen einsetzen zu können.

Unterschiede im Expertisegrad. Die gemittelten AvgLoss-Werte nehmen über die einzelnen Stichproben hinweg kontinuierlich ab: WB1 zeigte einen gemittelten AvgLoss-Wert von 0.471 ($SD = 0.090$), gefolgt von WB2 mit einem AvgLoss von 0.416 ($SD = 0.095$), WB3 mit einem AvgLoss von 0.365 ($SD = 0.004$) und WB4 mit 0.317 ($SD = 0.065$). Diese Abnahme wird auf dem 5%-Niveau signifikant (Kruskal-Wallis: $\chi^2(3, 26) = 7.893$, $p < .05$). Zur Verdeutlichung niedriger und hoher struktureller Unterschiede zwischen Lernerkarten und der Normkarte finden sich in Abbildung 61 je ein Beispiel eines Auszubildenden aus WB1 und aus WB4.

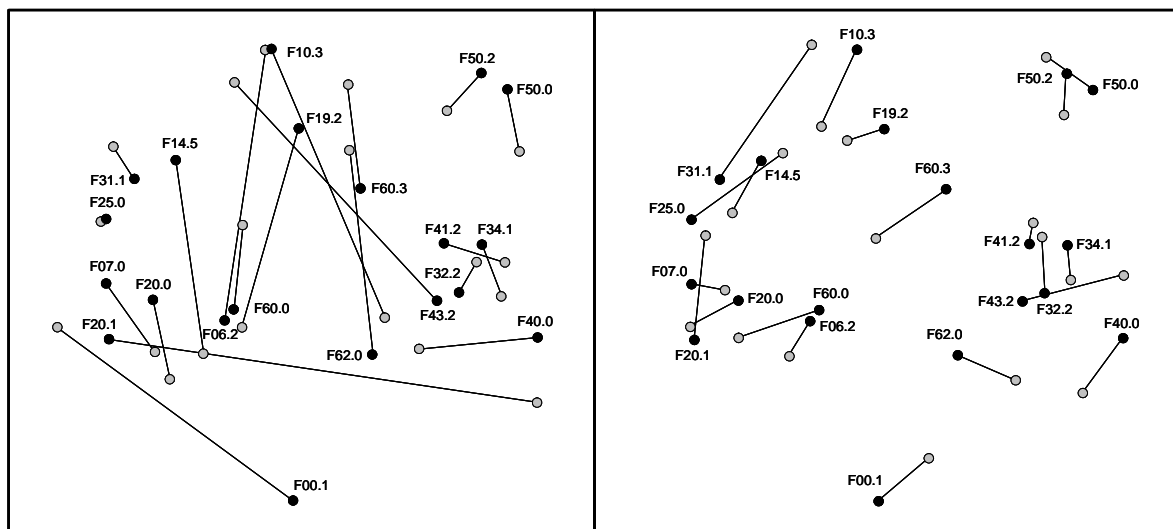


Abbildung 61. Zwei Prokrustes-Vergleiche individueller Karten mit dem Normmodell. Links Versuchsperson WB1-4, rechts Versuchsperson WB4-1. Die Abweichungswerte betragen für WB1-4 AvgLoss = 0.61, für WB4-1 AvgLoss = 0.29. Schwarze Punkte stehen für die Objektpositionen im Normmodell, graue für diejenigen der Lernenden.

Person WB1-4 weist in ihrer Wissenskarte hohe strukturelle Unterschiede zum Expertenmodell auf. Nur vereinzelte Störungsbilder befinden sich in der korrekten Region (F25.0, F31.1, F32.2, F34.1, F41.2, F50.0, F50.2). Mehrere von ihnen verletzen die Clusterzugehörigkeit. Die kognitive Struktur der psychopathologischen Störungsbilder hat sich also noch nicht adäquat ausgebildet. Im Gegensatz dazu besitzt Versuchsperson WB4-1 bereits eine sehr gute Vorstellung bezüglich der relationalen Beziehungen zwischen den Störungsbildern. Die Clusterzugehörigkeit der Objekte wird nur von ca. drei Störungsbildern verletzt (F20.1, F25.0, F31.1). Die geringen Unterschiede lassen auf ein hohes Verständnis in die Zusammenhänge und Unterschiede schließen.

Homogenisierung der Wissenskarten. Neben dem Vergleich der einzelnen Personen mit dem Normmodell ist vor allem auch interessant, ob es im Verlauf der Weiterbildung zu einer Homogenisierung der Sichtweisen zwischen den Personen kommt. Zu erwarten ist ja nicht nur, dass sich die individuellen Wissenskarten mit zunehmendem Expertisegrad dem Normmodell annähern, sondern dass sich infolge des gemeinsam besuchten Curriculums auch die strukturellen Sichtweisen untereinander systematisch ähnlicher werden. Um das zu überprüfen, werden alle individuellen Karten durch Prokrustes-Transformation paarweise miteinander verglichen. Die daraus generierte Matrix der AvgLoss-Werte zwischen je zwei individuellen Karten werden wiederum mit RobuScal skaliert und als Personenkarte dargestellt (sog. Prokrustesbasierte Personenskalierung: ProperScal; Läge, 2001). Abbildung 62 zeigt die zweidimensionale Darstellung der relationalen Beziehungen zwischen den Personenkarten.

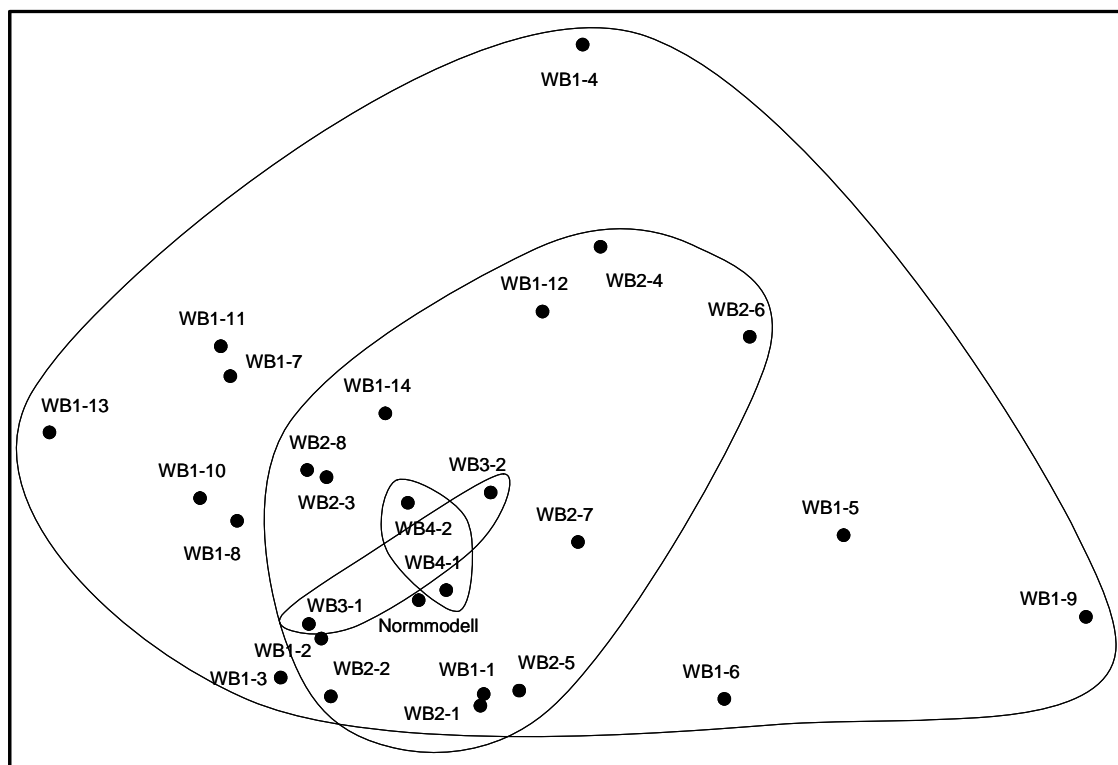


Abbildung 62. NMDS-Lösung aller individuellen Kartenpaarvergleiche (ProperScal). Die einzelnen Markierungen umfassen je die Teilnehmer der vier Stichproben (WB1-1 bis WB1-14; WB2-1 bis WB2-8; WB3-1 bis WB3-2; WB4-1 bis WB4-2). Ebenfalls eingetragen ist das Normmodell. Mittlerer AvgLoss aller Vergleiche = 0.52; Minimum AvgLoss = 0.27; Maximum AvgLoss = 0.80.

Deutlich zu erkennen ist die große Streuung der Kandidaten aus WB1. Mit zunehmender Weiterbildungszeit verringert sich diese Heterogenität. Alle vier Personen aus WB3 und WB4 liegen schließlich recht nahe beieinander (und auch am Normmodell).

8.4. Diskussion

Diese Arbeit stellt einen innovativen Ansatz zur Messung individuellen strukturellen Sachwissens im Bereich psychopathologischer Störungsbilder vor. Die Verfahrenslogik der Visualisierung struktureller Wissenscharakteristika einer Person über die Nonmetrische Multidimensionale Skalierung und der Performanzabschätzung über den Vergleich einer Lernerkarte mit einem Expertenmodell (Prokrustes-Transformation) soll einen Beitrag zur Standardisierung und Objektivierung von Methoden liefern, welche im Zusammenhang mit der Evaluation der Ergebnisqualität eines psychotherapeutischen Curriculums stehen. Das Verfahren trägt den Namen „Zürcher Wissensstrukturmodell für Psychopathologie (ZWP)“.

Dass die NMDS merkmalsbasiertes Sachwissen zu messen und darzustellen sowie die Prokrustes-Transformation Wissensveränderungen über einen Lernverlauf aufzuzeigen in der Lage ist, konnte bereits in mehreren kontrollierten Lernexperimenten gezeigt werden (vgl. Kapitel 2-4 in dieser Arbeit). In dieser Studie nun wurde nicht nur der Versuch unternommen, dieses Vorgehen auf das Inhaltsgebiet der psychopathologischen Störungsbilder auszuweiten, sondern zudem die Leistung der Methode im erstmaligen Einsatz in einem realen curricularen Umfeld auszuloten. Gezeigt werden konnten systematische strukturelle Wissensunterschiede zwischen den einzelnen Ausbildungsjahrgängen, und zwar ganz in der prognostizierten Abhängigkeit vom Expertisegrad. Außerdem ließ sich die Hypothese bestätigen, dass zunehmende Ausbildungszeit zu einer Homogenisierung der individuellen Sichtweisen führt. Dies bedeutet eine hinreichende Verfahrensleistung zur Wissensevaluation und ermöglicht den Aufbau von ZWP für den Praxiseinsatz als standardisiertes Messinstrument.

Die Messung der Ergebnisqualität eines Curriculums mit Hilfe von ZWP geschieht auf zwei Ebenen. Unter der Voraussetzung eines validen Vergleichskriteriums (Expertenmodell) können zum einen interindividuelle Stichprobenvergleiche helfen, den augenblicklichen Stand eines Ausbildungsjahrgangs festzuhalten. Schneidet beispielsweise ein fortgeschrittener Jahrgang im Mittel schlechter ab als ein jüngerer, so sollte entsprechend nach Ursachen gesucht und darauf reagiert werden. Sind die Wissenskarten einzelner Personen besonders auffällig, dann sind spezielle Förderungsmassnahmen im Aufbau des deklarativen Wissens angezeigt.

In einer Erweiterung der hier vorgestellten Auswertung der individuellen Wissenskarten kann über die Analyse der einzelnen Objektabweichungen außerdem festgestellt werden, mit welchen Störungsbildern die betreffende Person besondere Schwierigkeiten bei der relationalen Einordnung hat. Wir evaluierten die Verfahrenslogik von ZWP auf genau dieser Ebene (vgl. auch Kapitel 4 in dieser Arbeit). Nicht allein das Gesamtmaß AvgLoss, sondern die einzelnen Abweichungen der Objektpositionen einer Lernerkarte (ObjLoss) wurden auf ihre Vorhersagegenauigkeit bezüglich der Wissensqualität hin untersucht. Unter der Voraussetzung eines gewissen Wissensniveaus ist die Prokrustes-Transformation sogar in der Lage, auf Objektbene Wissensmängel (Lücken, Falschwissen) zu detektieren. Auf der Basis dieses Vergleichs

könnte neben der verstärkten Behandlung einzelner schwieriger Störungsbilder im Plenum zudem adaptiv auf den Wissensstand eines Lerners eingegangen werden. Dies steigert die Effizienz eines Curriculums, weil somit Mehrfachlernen bereits gewussten Materials praktisch komplett vermieden werden kann, indem man den Lerner dort abholt, wo er sich gerade befindet (vgl. Ausubel, 1963 und 2000; sowie zur Adaptivität von Lernprogrammen Leutner, 1992).

Die NMDS integriert ebenso einen didaktischen Aspekt. Über die Betrachtung eigener Wissenskarten und deren Abweichung zur Normkarte entsteht ein „externes“ Feedback über die individuellen Gedächtnisinhalte. Eine bewusste und aktive Beschäftigung mit den möglichen Ursachen von Kriteriumsabweichungen fördert den Lernprozess dahingehend, dass explizit nach Gemeinsamkeiten und Unterschieden gesucht wird und somit Stabilität und Differenziertheit des eigenen Wissens stark zunimmt. Dies kann sowohl in Eigenregie (selbstgesteuert) als auch gemeinsam mit Dozenten oder Kommilitonen geschehen. Die Ausbildung komplexer struktureller Facetten in dem für alles weitere fundamentale deklarative Wissen führt zu einer optimalen Vorbereitung des angehenden Therapeuten auf die zu erlernende praktische Tätigkeit.

Der Einsatz der NMDS zum Zwecke der Wissensmessung stellt unserer Ansicht nach einen innovativen und bislang einzigartigen Versuch dar, einen genau spezifizierten Teil der Ergebnisqualität eines psychotherapeutischen Curriculums über relationale Facetten von Gedächtnisinhalten zu überprüfen. Der im Vergleich zu herkömmlichen Prüfungsformen (z.B. Multiple-Choice, Forced-Choice) zusätzliche Nutzen dieses Verfahrens liegt in der Ausdrücklichkeit von Zusammenhängen. Expertenwissen zeichnet sich nicht allein durch eine große Wissensmenge aus, sondern lässt sich vielmehr anhand komplexer und differenzierter Wissenscharakteristiken erkennen (Gewichtungen, Prototypikalitäten, Auswahl der relevanten Beurteilungsdimensionen usw.). Das Bewusstsein bezüglich feiner Unterschiede und Gemeinsamkeiten ermöglicht die Grundlage für ein elaboriertes und adäquates Verhalten. Vor allem in einem sich in Entwicklung befindlichen und nicht kategorial strukturierten Inhaltsgebiet wie der Psychopathologie ist dies wichtig. Die sich vielfach überschneidende Phänomenologie mehrerer Störungen innerhalb eines Patienten benötigt ein gleichzeitiges Aufrechterhalten verschiedener Hypothesen. Nur dann ist der Therapeut in der Lage, das bestmögliche Therapieprogramm zu erstellen.

Wie bereits erwähnt, steht und fällt die Verfahrenslogik von ZWP mit der Qualität des Normmodells. Die Auswahl der zu beurteilenden Störungsbilder sollte dazu an die vermittelten Inhalte des Curriculums angepasst sein. Dieser Punkt ist sicherlich an der vorliegenden Pilotstudie zu bemängeln. Dass das zentrale Testkriterium der Annäherung der Lernerkarten an ein Normmodell mit zunehmender Expertise trotzdem bestätigt werden konnte, stimmt im Hinblick auf einen möglichen Praxiseinsatz von ZWP zuversichtlich. Es scheint möglich, damit ein zu den bereits bestehenden Verfahren der Evaluation komplementäres und vielseitig einsetzbares Evaluationswerkzeug für die relationale Komponente des deklarativen therapeutischen Wissens zu entwickeln.

8.5. Literatur

- Ausubel, D. P. (1963). *The psychology of meaningful verbal learning*. New York: Grune & Stratton.
- Ausubel, D. P. (2000). *The acquisition and retention of knowledge: A cognitive view*. Dordrecht: Kluwer Academic Publishers.
- Beitman, B. D. & Yue, D. (1999). A new psychotherapy training program: Description and preliminary results. *Academic Psychiatry*, 23 (2), 95-102.
- Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling – Theory and applications*. New York: Springer.
- Buchkremer, G. & Schneider, F. (1995). Qualitätssicherung in der Psychotherapie. In H.-J. Haug & R.-D. Stieglitz (Hrsg.), *Qualitätssicherung in der Psychiatrie* (S. 141-150). Stuttgart: Ferdinand Enke Verlag.
- Eckert, H. & Westhoff, K. (2000). Psychologiedidaktik und Evaluation am Beispiel der universitären Ausbildung im Fach "Diagnostik und Intervention". In G. Krampen & H. Zayer (Hrsg.), *Psychologiedidaktik und Evaluation II. Neue Medien, Psychologiedidaktik und Evaluation in der psychologischen Haupt- und Nebenfachausbildung* (S. 143-158). Bonn: Deutscher Psychologen Verlag GmbH.
- Egli, S., Schlatter, K., Streule, R. & Läge, D. (2006). A structure based expert model of the ICD-10 mental disorders. *Psychopathology* 39 (1), 1-9.
- Ehlert, U. & Kaiser, N. (2001). Curriculum der Ausbildung in Verhaltensmedizin. *Praxis Klinische Verhaltensmedizin und Rehabilitation*, 55, 199-202.
- Elizur, A., Kretsch, R., Spaizer, N. & Sorek, Y. (1994). Self-evaluation of psychotherapeutic competence. *British Journal of Medical Psychology*, 67 (3), 231-235.
- Foulkes, P. (2003). Trainee perceptions of teaching of different psychotherapies. *Australasian Psychiatry*, 11 (2), 209-214.
- Freedheim, D. K. & Overholser, J. C. (1998). Training in psychotherapy during graduate school. *Psychotherapy in Private Practice*, 17 (1), 3-18.
- Gigerenzer, G. (1981). *Messung und Modellbildung in der Psychologie*. München: Reinhardt.
- Gower, J. C. & Dijksterhuis, G. B. (2004). *Procrustes Problems*. New York: Oxford University Press.
- Jagsch, R., Kryspin-Exner, I. & Jandl-Jager, E. (2002). Selbsteinschätzung der Kompetenzen von Absolventen des psychotherapeutischen Propädeutikums. *Psychotherapie-Forum*, 10 (4), 214-222.
- Kaiser, N. & Ehlert, U. (2003). Weiterbildungsmöglichkeiten in der Verhaltensmedizin. *Psychotherapeut*, 48, 361-367.
- Läge, D. (2001). *Ähnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Läge, D., Daub, S., Bosia, L., Jäger, C. & Ryf, S. (2005). *Die Behandlung ausreißer-behafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag*. AKZ-Forschungsbericht Nr. 21. Zürich: Angewandte Kognitionspsychologie.
- Leutner, D. (1992). *Adaptive Lernsysteme. Instruktionspsychologisches Grundlagen und experimentelle Analysen*. Weinheim: Psychologie Verlags Union.
- Mandl, H. & Fischer, F. (Hrsg.). (2000). *Wissen sichtbar machen. Wissensmanagement mit Mapping-Techniken*. Göttingen: Hogrefe.
- Milne, D. L., Baker, C., Blackburn, I.-M., James, I. & Reichelt, K. (1999). Effectiveness of cognitive therapy training. *Journal of Behavior Therapy and Experimental Psychiatry*, 30, 81-92.
- Piaget, J. (1976). *Die Äquilibration der kognitiven Strukturen*. Stuttgart: Klett.
- Reinecker, H. & Schindler, L. (2000). Aus- und Weiterbildung. In J. Margraf (Hrsg.), *Lehrbuch der Verhaltenstherapie – Band 1* (2. vollst. über. & erw. Aufl.) (S. 655-662). Berlin: Springer.
- Schulz, W. (2000). Deutlich mehr Kompetenz. Erste Ergebnisse der Dokumentation und Evaluation der „Weiterbildung zum Sozialtherapeuten klientenzentriert/gesprächspsychotherapeutisch orientiert“. *Gesprächspsychotherapie und Personenzentrierte Beratung*, 31 (2), 146-151.

- Schulz, W., Flotho, B., Lübke, A. & Hahlweg, K. (1998). Dokumentation und Evaluation des Weiterbildungsganges Braunschweig in Klinischer Psychologie/Psychotherapie. *Verhaltenstherapie*, 8 (4), 234-244.
- Willutzki, U. & Botermans, J.-F. (1997). Ausbildung in Psychotherapie in Deutschland und der Schweiz und ihre Bedeutung für die therapeutische Kompetenz. *Psychotherapeut*, 42, 282-289.

9.

Pilot-INTUS

Zusammenfassung

Computergestützte, Adaptive Tutorielle Systeme (ATS) ermöglichen es, automatisiert auf den individuellen Wissensstand Rücksicht zu nehmen, indem der Lernprozess inhaltlich adaptiv gestaltet wird und Wissensdefizite so effizienter behoben werden können. Eine valide, vollständige und effiziente Wissensdiagnostik bilden dazu die Grundlage. Mithilfe der Nonmetrischen Multidimensionalen Skalierung (NMDS) und der Prokrustes-Transformation kann umfassend deklaratives Faktenwissen in Kognitiven (Wissens-) Karten dargestellt werden. Mehrere Studien haben die Tauglichkeit dieses Verfahrens bestätigt. Die Methode entdeckt unter bestimmten Voraussetzungen spezifische Wissensmängel und benötigt im Vergleich zur gesamten Informationsmenge verhältnismäßig wenig Zeit zur Diagnostik. Gewisse methodische Rahmenbedingungen führen zu einem didaktischen Konzept, welches ausgehend von einem Basiswissen neue Elemente in assimilativer Form mit bestehenden Gedächtnisinhalten verknüpft. Bislang beschränkten sich die Untersuchungen aber auf die Beantwortung einzelner Fragestellungen. In einem komplex strukturierten Lernexperiment sollen nun erstmals die bekannten Erkenntnisse bezüglich der Wissensdiagnostik und das gestuft modulare didaktische Konzept umfassend geprüft werden. Im Fokus der Aufmerksamkeit dieses Experimentalkonzeptes stehen die Replikation methodischer Performanz und die Überprüfung der Tauglichkeit der Adaptivitätslogik. Trotz bestätigten methodischen Erwartungen und tendenziellen Treatment-Effekten konnte keine der zentralen Hypothesen bestätigt werden und das Experiment muss insgesamt als gescheitert betrachtet werden. Mögliche Ursachen und Empfehlungen für eine weitere Wiederholungsstudie werden diskutiert.

Schlüsselwörter: Kognitive Struktur, Kognitive Karte, Nonmetrische Multidimensionale Skalierung (NMDS), Prokrustes-Transformation, Adaptives Tutorielles System (ATS), Wissenserwerb, Wissensdiagnostik, Adaptive Lernempfehlungen

9.1. Einleitung

Adaptive Tutorielle Systeme. E-Learning, Computer Based Training (CBT), Computer Assisted Instruction (CAI), Web Based Training (WBT) oder Virtuelle Lernprogramme sind, um nur einige wenige Ausdrücke zu nennen, vielfältige Bezeichnungen für die Verwendung des Computers als Lern- und Lehrmedium, welche im Prinzip und in seiner allgemeinen Form alle das Gleiche meinen: Mittels eines Rechners greift ein Lerner auf multimedial aufbereitete Lernmaterialien statischer und interaktiver Natur zu. Damit unterscheidet sich das Lernen am Computer eigentlich nicht grundsätzlich von jenem mit einem Lehrbuch, lässt man die sich neu ergebenden Möglichkeiten solch eines Systems außer Acht. Ein großer und prominenter Vorteil liegt in der Unabhängigkeit von Ort und Zeit des Lernprozesses. Der heutige gesellschaftliche Lebensstil erwartet ein hohes Maß an Flexibilität im Sinne einer stetigen Weiterbildung und andauernden Neuorientierung sowie beschränkter zeitlicher Ressourcen, die optimal genutzt werden sollen. Die Pluralität des geforderten Verhaltens eines Subjekts bspw. im Zuge der Bologna-Reform erfordert nun auch eine Anpassung und Innovation der Aus-

Weiter- und Fortbildung. Der Einsatz von Computern in den verschiedensten Bildungsinstitutionen kann eine Erleichterung für diesen Prozess sein, schreitet mit großen Schritten voran und führt den seit rund einem Jahrzehnt in größerem Umfang stattfindenden Wandel vor Augen.

Ein zweiter wichtiger Vorteil von virtuellen Lernprogrammen ergibt sich dann, wenn nicht nur Informationen passiv auf dem Bildschirm präsentiert werden, sondern beim Versuch, das System mit einer gewissen „Intelligenz“ zu versehen. Der klassische Frontalunterricht (Face-to-Face) bspw. kennzeichnet sich durch einen Dozierenden aus, welchem viele Auszubildende gegenüberstehen. Trotz der großen Masse besteht dort natürlich die Möglichkeit, studentische Fragen zu beantworten um Unklarheiten aus dem Weg zu räumen. Dies funktioniert aber nur aus dem Grund, da der Referent über Expertise verfügt und flexibel seinen Unterricht an die situativen Gegebenheiten anpassen kann, indem er z.B. ein bereits behandeltes Thema erneut und vertiefend vermittelt oder bereits Bekanntes überspringen kann. Unbestritten ist aber der Rahmen solcher flexiblen Handlungsmöglichkeiten begrenzt, da allen Anwesenden der Lehrstoff in einer gewissen Zeit beigebracht werden muss und damit nicht genügend Raum bleibt, auf alle individuellen Stärken und Schwächen Rücksicht zu nehmen. Ein Computerprogramm kann dagegen auf den jeweiligen Benutzer eingehen und analog einer menschlichen Lehrkraft individualisierte Anpassungen vornehmen und bereitstellen. Solche Wünsche, den Computer als intelligente Hilfskraft nutzbar zu machen sind nicht neu und mündeten nach anfänglichen, mehr oder weniger passiven Konzepten von automatisierten Lehrmaschinen und praktischen Umsetzungen (vgl. Pressey, 1926; Skinner, 1958; Crowder, 1959) in solchen, die auch auf instruktionalistischer Seite sich mehr und mehr dem menschlichen Tutor anzugleichen versuchten. Genannt werden sie Intelligente Tutorielle Systeme (kurz: ITS; vgl. Goldstein, 1982 und zu den Komponenten eines ITS Van Merriënboer, Van Den Berg & Maaswinkel, 1989) oder auch Adaptive Tutorielle Systeme (Leutner, 1992).

Solche Systeme verfügen über Lehrstoffexpertise und ermöglichen dadurch die inhaltliche Anpassung des Systems an die spezifischen Charakteristika des Einzelnen. Zeigt nun eine Person gewisse Defizite in ihrem Wissensstand, bietet das Programm in analoger Form zu einer Lehrkraft verstärkt diejenigen Inhalte an, welche noch mangelhaft im Gedächtnis gespeichert sind. Diese Adaptivität führt zu einer nicht-linearen Lern-Struktur und verhilft zu größerer Wissensqualität und Ökonomie, da lediglich auf denjenigen Inhalt fokussiert wird, der den größten Lerngewinn verspricht (vgl. zum Thema Adaptive Lernprogramme Leutner, 1992a, 1995). Die Gewährleistung inhaltlicher Adaptivität benötigt eine kontinuierliche Statusdiagnose des momentanen Lernerwissens. Diese muss umfassend ausgerichtet sein, zum einen aus dem Grund, dass jede Lücke entdeckt werden kann und zum anderen, weil ein vom Lerner selbstgesteuerter Kontrollprozess meist mangelhaft ist (Charney, Reder & Kusbit, 1990; Strohschneider, 1990), nicht zuletzt darum, weil Falschwissen subjektiv nicht als Defizit erkannt wird. Der Diagnostikaufwand kann sehr unterschiedlich ausfallen, je nachdem welche Wissenskategorie getestet wird. Aus der Forschung zur Diagnostik von Regelwissen ist bspw. bekannt, dass für das Verständnis der Wahrscheinlichkeitsrechnung nur einige wenige Regeln für das Gesamtverständnis beherrscht werden müssen (vgl. Albert & Lukas, 1999 oder für eine Replik Läge, 2001). Bei der Überprüfung von deklarativem Sachwissen müsste

für eine vollständige Befragung jedes Faktum einzeln einem Test unterzogen werden. In der Praxis ist dies ein nicht durchführbares Unterfangen, da die notwendige Zeit zur Diagnose den gesamten Lernprozess wieder ineffizient werden lässt (vgl. Leutner, 1992b).

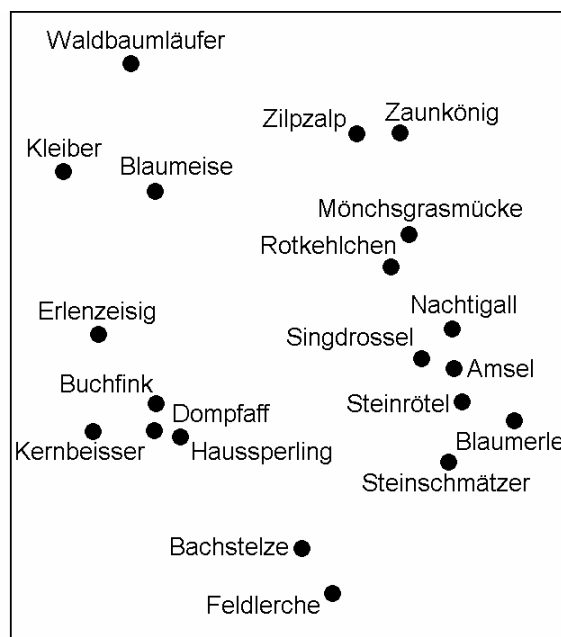


Abbildung 63. NMDS-Lösung (Kognitive Karte), basierend auf Ähnlichkeitsbeurteilungen (Expertenurteile), welche im Experiment als Normmodell Verwendung findet.

Wissensdiagnostik. Ein möglicher Ansatz, für merkmalsbasiertes Sachwissen all die Fakten relativ schnell einem Test zu unterziehen, bietet die so genannten „Kognitive Karte“. Die Nonmetrische Multidimensionale Skalierung (NMDS) – ein Verfahren aus der Familie der multivariaten Analysemethoden – stellt assoziative, objektbasierte Daten in einem mehrdimensionalen, euklidischen Raum dar, innerhalb dessen sehr ähnliche Elemente nahe beieinander und unähnliche weit voneinander weg positioniert sind (s. Abbildung 63; vgl. zur Logik des Verfahrens Borg & Groenen, 1997; Läge, Daub, Bosia, Jäger & Ryf, 2005). Die Rohdaten entstammen meist aus Ähnlichkeitsbeurteilungen zwischen einem Satz von Objekten. Die Berechtigung, Ähnlichkeiten als Grundlage zur Wissensmessung zu verwenden, liegt darin, dass wir Menschen bereits mit wenigen Informationen in der Lage sind, ein Urteil über zwei Dinge abzugeben und dieses Urteil auf der Basis von relevant erscheinenden Attributen zu bilden (Klauer, 1989). Die Bildung eines Ähnlichkeitswertes verstehen wir als einen mehr oder weniger automatisiert ablaufenden Prozess, innerhalb dessen verschiedene Merkmale und deren Ausprägungsunterschiede zur Beurteilung der Relationalität zwischen Objekten herangezogen werden. Damit geht diese Logik der Wissensmessung über das reine Abfragen von Merkmalslisten hinaus, da zusätzlich zur Merkmalsinformation auch relationale Gefüge, Integration und Elaboration (Gewichtungen, kategoriale Strukturierung usw.) der Wissensstruktur mit gemessen werden. Zusätzlich sind sie unabhängig von Frageformaten und relationalen Vorschriften. Ein differenzierteres und umfangreicheres Merkmalswissen zeigt sich in einer qualitativ besseren Ähnlichkeitseinschätzung und in der Folge in einer im Vergleich zu einem („objektiv“ festgelegten) Zielmodell optimaleren räumlichen Positionierung des Ob-

jekts in der NMDS-Karte, womit die implizit vorhandene Merkmalsinformation einer weiteren quantitativen Beurteilung zugänglich gemacht werden kann.

Kognitive Karten können über die Prokrustes-Transformation (Gower & Dijksterhuis, 2004) auf Übereinstimmung bzw. Abweichung geprüft werden (quantifiziert über den so genannten Average Loss oder AvgLoss: vgl. Läge, 2001). Dabei wird die Testkarte auf die Zielkarte (z.B. das Expertenmodell) gelegt und über die vier distanzinvarianten Transformationsregeln Drehung, Spiegelung, Schiebung und Skalierung (Streckung, Stauchung) die bestmögliche Passung bestimmt (Abbildung 64). Die verbleibenden großen einzelnen Objektabweichungen (quantifiziert über den Object Loss oder ObjLoss) deuten auf mangelhaftes Merkmalswissen und würden innerhalb eines ITS/ATS vom Verfahren detektiert und für die Zusammenstellung der folgenden Lektion als Erste ausgewählt werden. Mit dieser kurz vorgestellten Verfahrenslogik besteht ein möglicher Ansatz, auf der einen Seite die notwendige Leistung einer umfassenden Wissensdiagnostik zu erbringen und zum anderen so effizient wie möglich zu sein. Noch nicht abschließend geklärt ist aber die Performanz der NMDS und Prokrustes-Transformation als Wissensmessungsinstrument innerhalb eines möglichst realen und gesamthaft konzipierten Lernsettings.

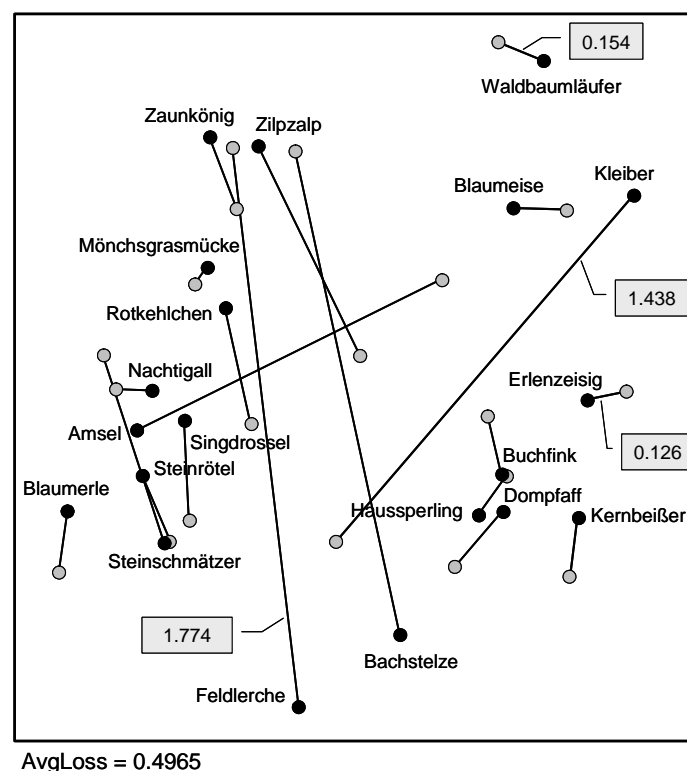


Abbildung 64. Ergebnis einer Prokrustes-Transformation. Die schwarzen Punkte stehen für die Objektpositionen im Expertenmodell, die grauen für die korrespondierenden Positionen eines (fiktiven) Lerner. Vier Beispiele von quantifizierten Objektabweichungen (ObjLoss) sind numerisch eingetragen. Je höher der Wert, desto größer die Abweichung, welche auf mangelhaftes Wissen über das jeweilige Objekt hinweist. Die mittlere Abweichung aller Objekte ergibt das Gesamtmaß AvgLoss (hier: 0.4965).

Leistungen und Grenzen der NMDS. Betrachten wir im Folgenden die Erkenntnisse aus mehreren abgeschlossenen Studien, die sowohl simulativ als auch experimentell die Tauglichkeit

des Verfahrens als Instrument zur Sachwissensmessung hervorheben, nichtsdestotrotz aber aufgrund gewisser spezifizierender Rahmenbedingungen grundsätzliche Implikationen für die praktische Anwendung nach sich ziehen.

Es ist möglich, mit dem Verfahren der NMDS deklaratives Faktenwissen verschiedensten Inhalts in Kognitiven Karten darzustellen (z.B. Egli, Schlatter, Streule & Läge, 2006; Läge, 2001, Marx & Hejj, 1989; Marx & Läge, 1995). Außerdem können Lernprozesse auf einer globalen Ebene verfolgt werden. Der Lernfortschritt im Merkmalswissen führt zu einer Kongruenz der individuellen Kognitiven Karten in die Richtung des Normmodells und einer Homogenisierung der zu Beginn des Wissenserwerbs idiosynkratischen NMDS-Lösungen (vgl. Kapitel 2, 5 und 8 in dieser Arbeit). Darüber hinaus ist es möglich, Vorhersagen über die Wissensqualität einzelner Objekte zu machen. Anhand der Höhe der relativen Objektabweichung „ObjLoss“ sind mangelhaft gespeicherte Objekte überzufällig weiter von der Position im Normmodell weg verortet als gut gewusste Objekte (vgl. Kapitel 4 in dieser Arbeit). Allerdings nur unter bestimmten Voraussetzungen: Zum einen ergibt sich lediglich dann eine akzeptable Vorhersagegenauigkeit, wenn die zu testende Objektzahl ≥ 16 beträgt. Ab einer Objektzahl von 20 verbessert sich die Diagnoseleistung nicht mehr evident und wird als minimale Objektmenge zur Diagnostik von Sachwissen über die NMDS empfohlen (vgl. Kapitel 4 in dieser Arbeit). Zum anderen müssen die Lerner über ein bestimmtes Maß an Vorwissen verfügen. In simulativen Studien mit dem Inhaltsgebiet Geographie ergab sich ein für eine akzeptable Diagnosegenauigkeit notwendiger Wert von rund 60% korrekt gewussten Objekten, in einem empirischen Experiment mit dem komplexeren Inhaltsgebiet Ornithologie von rund 80%. Der Grund für diese doch beträchtliche Forderung nach Vorwissen liegt in der stabilisierenden Kraft von zueinander passenden Ähnlichkeitsurteilen, welche auf der Basis von gewusstem und differenziertem Merkmalswissen gebildet werden können. Diese Urteile sind zueinander konsistenter und transitiver als Urteile zwischen Objekten, über die die Person nur wenig weiß. Konsistente Urteile werden vom hier verwendeten RobuScal-Algorithmus (Läge, Daub, Bosia, Jäger & Ryf, 2005) prioritär bei der räumlichen Positionierung behandelt und platzieren sich dementsprechend auch valider im Vergleich zum Expertenmodell als die nicht gewussten Objekte.

Kurz angesprochen wurde bereits das Problem der umfassenden und demzufolge aufwändigen Diagnostik, wenn man maximal adaptiv sein will. In der klassischen Form müssen alle Paarvergleiche zwischen den interessierenden Elementen des Wissensgebietes erhoben werden (Die resultierende Anzahl ergibt sich über die Formel $n * (n - 1) / 2$). Bei 20 Objekten müssen bspw. 190, bei 40 Objekten bereits 780 Ähnlichkeiten beurteilt werden (letztere Größe ist in der Praxis ein nicht zumutbares Unterfangen)¹⁵! Die von Leutner (1992b) allgemein als „Testlängendilemma“ bezeichnete Ineffizienz solch einer umfassenden Wissensdiagnostik entschärft sich aber bei der Verwendung der NMDS als Diagnoseinstrument über die Implemen-

¹⁵ Erfahrungsgemäss benötigt die Einschätzung von 190 Urteilen bereits zwischen 15 und 35 Minuten, wobei zum einen eine grosse personenspezifische Varianz festzustellen ist und zum anderen Individuen mit der Zeit und Vertrautheit mit dem Verfahren erheblich an Schnelligkeit zulegen.

tierung einer gestuften didaktischen Ausrichtung des Lernprozesses erheblich: Modelle über den Wissenserwerb besagen, dass, ausgehend von einem stabilen und qualitativ guten Überblickswissen, neu hinzukommendes Lernmaterial in assimilativer Weise mit bestehenden Gedächtnisinhalten verknüpft wird und optimalerweise über einen förderlichen didaktischen Aufbau auch so verknüpft werden soll (vgl. Ausubel, 2000; Piaget, 1976). Verfügt eine Person über einen für das Inhaltsgebiet repräsentativen und adäquat strukturierten Ausschnitt aus dem Wissensfeld, können neue Fakten ohne großen kognitiven Aufwand und vor allem ohne die Notwendigkeit einer Umstrukturierung der bereits vorhandenen Struktur (dies käme einer Akkomodation gleich) integriert werden. Diese wichtige lerntheoretische Erkenntnis erlaubt es, vormals umfassend und stabil Gelerntes als gegeben zu betrachten und demzufolge ist eine kontinuierliche Überprüfung dieser Elemente unnötig. Die Wissensdiagnostik neu gelernter Materials kann sich lediglich auf die neu gebildeten Gedächtnisinhalte beschränken. Dass Wissen im Langzeitgedächtnis mit einer differenzierten Verarbeitungstiefe und gemessen mittels der NMDS unwesentlichen Veränderungen unterliegt, konnte in einem Lernexperiment mit Messwiederholungen (vgl. Kapitel 5 in dieser Arbeit) gezeigt werden. Intraindividuelle Kartenvergleiche aus Messwiederholungen und bei kontrolliert gleichem Wissensstand ergaben ausschließlich geringfügige und lediglich personenspezifische, unsystematische Unterschiede im AvgLoss-Wert ($\text{AvgLoss} = 0.17$), was als Rauschen im Urteilsverhalten bezeichnet und somit vernachlässigt werden kann. Dies ist eine notwendige Voraussetzung, wenn im weiteren Verlauf des Lernprozesses nur noch Ähnlichkeitsurteile zwischen gerade erst gelerntem und den Elementen des Vorwissens erhoben werden möchten. Der vormals exponentielle Zusammenhang zwischen Urteils- und Objektmenge reduziert sich dadurch in einen linearen und führt zu einem mit steigender Objektzahl immer größer werdenden Effizienzgewinn.

Wie bereits erwähnt, führen die Resultate aus diesen, sich auf spezifische Fragestellungen beschränkenden Studien zu formalen Bedingungen, die bei einem umfassenden praktischen Einsatz der Verfahrenslogik zu beachten sind. Über ein differenziertes Lernsetting soll nun ein erstes Pilot-Experiment konzipiert, in welchem Laien den gesamten Lernprozess durchlaufen und auf ein expertennahes Niveau gebracht werden. Auf der einen Seite ergibt sich dadurch die erstmalige Überprüfung der Tauglichkeit des gestuften, modularen Lernprozesses als Gesamtes. Auf der anderen Seite erlaubt dies die Replikation bestehender Resultate im Zusammenhang mit methodischen Fragestellungen.

Erwartete Befunde. Betrachten wir im Folgenden den detaillierten Lernprozess, die Operationalisierung der Konzepte und die erwarteten Befunde. Das Lernen beginnt mit dem Aufbau einer Grundstruktur, welche den Kognitiven Raum „aufspannt“, den Lernern also ermöglicht, die Dimensionalität und kognitive Gewichtung zu erkennen, welche wichtig ist, um dann in einem zweiten Schritt dieses Wissen assimilativ zu erweitern. Diese Grundstruktur (GS) soll 20 Elemente umfassen, damit die Kombination von NMDS und Prokrustes-Transformation als valides Diagnostikinstrument greift. Der besseren Kontrollierbarkeit wegen lernen die Versuchspersonen kriteriumsabhängig eine bestimmte minimale Menge an Objekten (60% oder 12 Objekte) aus der GS. Anschließend unterteilt sich die Stichprobe in eine Experimental- und eine Kontrollgruppe. Der Inhalt der Lektion für die Experimentalgruppe wird adaptiv

anhand des Kartenvergleichs (Prokrustes-Transformation) der individuellen Wissenskarte mit der Expertenkarte bestimmt (Adaptiv-Gruppe). Die Kontrollgruppe lernt demgegenüber eine vorgängig festgelegte Reihenfolge von Merkmalsinformationen (Standard-Gruppe). Die individualisierte adaptive Auswahl soll, so die Hypothese, effizienter im Sinne eines größeren Umfangs an Merkmalsinformation und einer qualitativ besseren kognitiven Struktur sein, da anhand der Prokrustes-Transformation Wissensdefizite entdeckt und somit spezifischer behoben werden können. Die Überprüfung der Vorhersage erfolgt über einen Test des Merkmalswissens (Punktzahl in einem Multiple-Choice-Fragebogen) und über die Abweichung der individuellen Kognitiven Karten zur Expertenkarte (AvgLoss).

Ein zweiter Schwerpunkt beschäftigt sich mit der Sensitivität von NMDS-Lösungen und Prokrustes-Transformation. Die Tauglichkeit dieses Vorgehens wird überprüft, indem über den Kartenvergleich der individuellen NMDS-Lösungen mit der Expertenkarte anhand der Objektabweichungen (ObjLoss) der neuen Objekte die Diskriminationsleistung des Verfahrens bestimmt werden kann. Objekte, welche gelernt wurden, sollen überzufällig niedrigere Abweichungen zur Expertenkarte aufweisen als noch nicht gelernte. Zwei verschiedene methodische Bedingungen sollen diesbezüglich untersucht werden. Erstens unter der „klassischen“ Bedingung der Skalierung von Proximitätsmaßen als Gesamtes. Zweitens bei einer Erweiterung des Grundwissens und einem reduzierten Diagnostikkonzept. Indem auf Urteile zwischen bereits intensiv und wiederholt Gelerntem verzichtet wird und so neue zu diagnostizierende Objekte methodisch in eine bestehende Struktur integriert werden, kann eine erhebliche Reduktion des Diagnostikaufwandes erreicht werden. Trotzdem darf die Treffergenauigkeit des Verfahrens nicht maßgeblich darunter leiden.

Sowohl die Vorhersagegenauigkeit der Kognitiven Karten in Bezug auf die Detektion mangelhaften Wissens (anhand des ObjLoss) als auch das Adaptivitätskonzept werden überprüft (anhand AvgLoss und Fragebogen).

9.2. Methode

Stichprobe. Lediglich sechs Personen (1 w, 5 m) konnten für die freiwillige Teilnahme an diesem sehr zeitaufwändigen Experiment gewonnen werden. Die Versuchspersonen verfügen über einen Hochschulabschluss oder sind Studierende der Universität Zürich. Je drei Versuchspersonen wurden zufällig der Adaptiv-Gruppe und der Standard-Gruppe zugeordnet¹⁶.

Lernmaterial. Eine repräsentative Auswahl aus der mitteleuropäischen Singvogelwelt bildet das zu lernende Wissensgebiet. Aussehen (als graphische Abbildung), Nahrung (als Text), das Habitat (als Text), Familienzugehörigkeit (als Text) und Verhalten (als Text) bilden den Wissensumfang und sind pro Vogel auf Kärtchen im Format A5 notiert. Der Wissenstest (Frage-

¹⁶ Aufgrund der geringen Anzahl an Probanden muss die Aussagekraft der Resultate als sehr gering eingestuft werden. Die Darstellung der Resultate beschränkt sich demzufolge auf eine zwar umfassende, aber eher als Leitfaden für eine valide Analyse zu verstehende Beschreibung des Gefundenen und ohne die Interpretationen der Resultate über die rein tendenzielle Schlussfolgerung zu heben!

bogen) befragt lediglich die ersten drei Merkmale; Familienzugehörigkeit und Verhalten dienen lediglich als Lernunterstützung im Sinne einer Differenzierung und Erleichterung der Strukturalisierung des Wissens aufgrund des entstehenden Gesamteindrucks. Die Beschränkung auf die drei Merkmale Aussehen, Habitat und Nahrung ergibt sich aus der Befragung von Experten, welche diese drei Attribute als zentrale Beschreibungsdimensionen für das Inhaltsgebiet Ornithologie befanden (Daub, 2001). Außerdem wurde darauf geachtet, dass die Objektauswahl auf diesen drei Merkmalsdimensionen möglichst breit streuen, d.h. dass der gesamte Werte- oder Ausprägungsbereich innerhalb der mitteleuropäischen Singvögel repräsentativ abgedeckt ist. Die Text- und Bildinformation stammt aus Ornithologie-Fachliteratur (Bezzel, 1993; Hudec, 1992; United Soft Media, 2001).

Experimental design. Die Versuchspersonen durchliefen vier Sitzungen im Abstand von zwei bis drei Tagen. Im Folgenden werden die Experimentalteile im Einzelnen beschrieben:

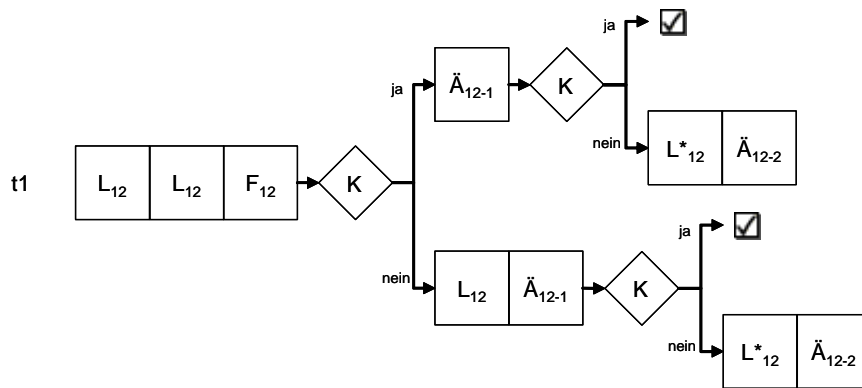


Abbildung 65. Experimentaldesign Zeitpunkt t1.

Die erste Sitzung t1 (Abbildung 65) dient dem Erwerb des Vorwissens (Grundstruktur, GS). In den einleitenden Worten haben wir gesehen, dass für eine valide Wissensdiagnostik mindestens 60% Vorwissen vorhanden sein sollte, damit über die Prokrustes-Transformation (ObjLoss) eine akzeptable Vorhersagegenauigkeit hinsichtlich mangelhaft gewusster Objekte erzielt werden kann. Zwei Lerneinheiten à 15 Minuten und über zwölf Objekte (L_{12}), unterbrochen durch die ausführliche Experimental-Instruktion, bilden den Start in das Experiment. Anhand eines Fragebogens (F_{12}) wird dann überprüft, ob das Kriterium (K) von 75% korrekt zugeordneter Merkmalsausprägungen erreicht ist (das Kriterium von 75% orientiert sich am Resultat einer experimentellen Studie zur Überprüfung der Güte des ObjLoss; vgl. Kapitel 4 in dieser Arbeit). Wenn ja, beurteilen die Versuchspersonen die Ähnlichkeiten zwischen diesen Objekten (\check{A}_{12-1}) und schließen die erste Sitzung ab, wenn der AvgLoss zur Normkarte den Wert von ≤ 0.3 erreicht (dieser Wert ergibt sich aus einem Lernexperiment, innerhalb dessen die Versuchspersonen auf ein expertennahes Niveau gebracht wurden; vgl. Kapitel 2 in dieser Arbeit). Bei Nichterfüllen des Fragebogen-Kriteriums folgt vor der Abgabe der ersten Ähnlichkeitsurteile eine Repetition des Merkmalswissens. Erreichen die Lerner keine genügend gute Kartenqualität, erhalten sie die Möglichkeit, das Merkmalswissen mithilfe der Normkarte als visuelle Strukturierungsunterstützung (L^*_{12}) erneut zu lernen, gefolgt von einer abschließenden Ähnlichkeitsbeurteilung (\check{A}_{12-2}).

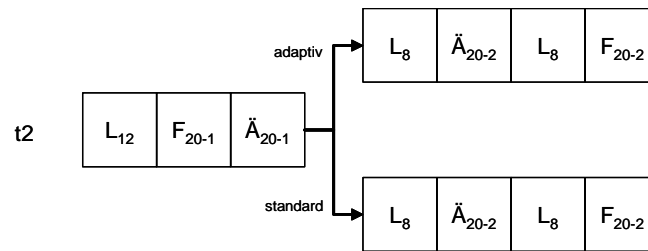


Abbildung 66. Experimentaldesign Zeitpunkt t2.

Mit der Repetition der bereits bekannten zwölf Elemente (L_{12}) begannen die Lerner die zweite Sitzung t2 (Abbildung 66). Der Fragebogen (F_{20-1}) und die Ähnlichkeitsurteile (\check{A}_{20-1}) bildeten die Basis, um die Probanden in homogene Gruppen aufzuteilen. Je die Hälfte der Personen wurde der *Adaptiv*-Gruppe bzw. der *Standard*-Gruppe zugeordnet. Die Adaptiv-Gruppe erhielt die acht zu lernenden Objekte (L_8) anhand der maximalen Objektabweichungen (ObjLoss) aus dem Kartenvergleich zwischen individueller und Normkarte (Lerndauer: 15 Minuten). Eine erneute Überprüfung des strukturellen Wissens (\check{A}_{20-2}) führte zur Auswahl der Inhalte der zweiten Lektion (anhand der maximalen ObjLoss-Werte), gefolgt von einer Überprüfung des deklarativen Merkmalwissens anhand eines Fragebogens (F_{20-2}). Der Experimentalablauf für die Standard-Gruppe war exakt gleich strukturiert mit der Ausnahme, dass für jede Versuchsperson und Lektion je zwei Zufallsziehungen von acht aus den 20 Elementen gebildet wurde.

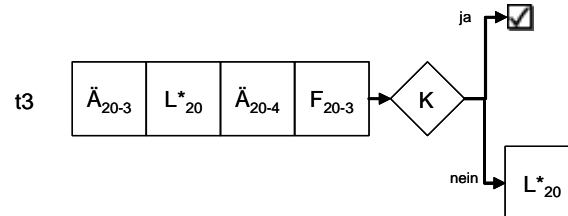


Abbildung 67. Experimentaldesign Zeitpunkt t3.

Die dritte Sitzung t3 (Abbildung 67) hatte zum Ziel, das Wissen der Lerner über die 20 Basisobjekte auf ein expertennahes Niveau zu bringen um in t3 von einer stabilen Grundstruktur für eine assimilative Erweiterung der Wissensfeldes ausgehen zu können. Die der ersten Ähnlichkeitsbeurteilung (\check{A}_{20-3}) folgende 15-minütige Lektion (L^*_{20}) umfasste das Merkmalswissen der 20 Objekte inklusive der Normkarte als visuelle Strukturierungslernhilfe. Eine erneute Ähnlichkeitsbeurteilung (\check{A}_{20-4}) und der Fragebogen (F_{20-3}) bildeten das Kriterium (K). Im Fragebogen mussten 75% der Antworten korrekt sein und die individuelle Kognitive Karte durfte einen AvgLoss-Wert von 0.3 zur Normkarte nicht überschreiten. Waren die Kriterien nicht erfüllt, folgte eine erneute Lernphase (L^*_{20}) mit der Normkarte als Lernhilfe von zehn Minuten mit speziellem Fokus auf diejenigen Objekte, welche im Kartenvergleich die höchste Positionsabweichung hatten (maximale ObjLoss-Werte).

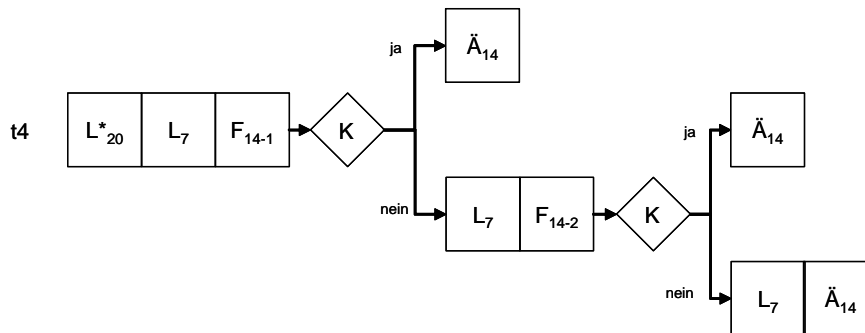


Abbildung 68. Experimentaldesign Zeitpunkt t4.

Die vierte Sitzung t4 (Abbildung 68) begann mit einer Repetition inklusive der Normkarte (L^*_{20}), gefolgt von einer Erweiterung des Grundwissens. Vp 1 erhielt eine Zufallsauswahl von sieben Objekten aus einem Pool von 14 neuen Vögeln (L_7). Vp 2 lernte die komplementären Objekte. Für Vp 3 wurde eine neue Zufallsauswahl gebildet, Vp 4 erhielt die komplementären Objekte usw. Nach der Lerndauer von 15 Minuten wurde das Wissen in einem Multiple-Choice Fragebogen (F_{14-1}) überprüft (Dieser beinhaltet jeweils auch die sieben noch nicht gelernten Vögel.). Ist das Kriterium von „6 von 7 Objekten“ richtig zugeordnet erfüllt, gaben die Probanden Ähnlichkeitsurteile (\ddot{A}_{14}) von den 14 neuen Objekten zu den 20 der Grundstruktur ab (= 280 Urteile). Auf die Urteile innerhalb der neuen und innerhalb der Grundstrukturelemente wurde verzichtet. Bei Nichterreichen des Kriteriums lernten die Versuchspersonen nochmals die sieben neuen Objekte und wurden wiederum getestet (F_{14-2}). Bei nochmaligem Nicht-Erreichen des Kriteriums folgte ein abschließender dritter Zyklus, bevor auf jeden Fall die ähnlichkeitsbasierte Beurteilung folgte.

Normmodell. Drei verschiedene Normmodelle werden benötigt. Diese Karten dienen als Vergleichskarte/Normkarte für die Prokrustes-Transformation, zwei davon auch als strukturierende Lernhilfe: Die 12er-Normkarte aus t1 und die 20er-Karte aus t2 und t3 sind sowohl Gütekriterium als auch didaktische Lernunterstützung (um die Strukturierungsleistung bei den Probanden zu unterstützen). Sechs Experten haben zu diesen 20 Vögeln Ähnlichkeitsurteile abgegeben (vgl. Kapitel 4 in dieser Arbeit). Die gemittelten Matrizen bilden die Datenbasis zur Berechnung der beiden Expertenmodelle (bestehend aus 12 und 20 Objekten). Für t4 wird eine Normkarte bestehend aus 34 Objekten verwendet. Weitere sechs Ornithologie-Experten haben in die vorhandene 20er-Karte (vgl. Kapitel 4) von Hand 18 neue Objekte eingetragen und so die Positionskoordinaten generiert. Diejenigen vier Vögel, welche von den Experten am unterschiedlichsten positioniert wurden, wurden aus der endgültigen Auswahl extrahiert. Die gemittelten Positionen der 14 neuen, kombiniert mit den Positionen der 20 Basis-Vögel ergaben das 34er-Normmodell, welches nur als Vergleichsmaß Verwendung findet.

Berechnung der individuellen Kognitiven Karten der Lerner. Die individuellen NMDS-Lösungen werden mittels des RobuScal-Algorithmus erstellt (Läge, Daub, Bosia, Jäger & Ryf, 2005). Dieser ist, im Gegensatz zu klassischen NMDS-Verfahren, robust gegen Ausreißer und Streuung in den Ähnlichkeitsurteilen und erlaubt die Interpretation der Kognitiven Karten auf individuellem Niveau. Für die 12er und die 20er Karten aus t1 bis t3 werden alle Paarvergleiche erhoben. Die Erstellung der 34er Karten aus t4 erfolgt über eine sequentielle

Einpassung der 14 neuen Objekte in die 20er Grundstruktur, da nicht eine komplette Urteils-matrix vorhanden ist, sondern lediglich Urteile zwischen den neuen und den Objekten der Grundstruktur zur Verfügung stehen. Um stabilere und vergleichbare Positionierungen der neuen Objekte zu erhalten, werden diese sequentiell in die 20er Expertenstruktur eingepasst (ausführlicher ist die Logik des Verfahrens in Kapitel 5 in dieser Arbeit beschrieben).

Fragebogen. Für das Erreichen der Wissenskriterien und als Kontrollüberprüfung fungieren Merkmals-Fragebögen. Das Aussehen der Objekte (Abbildung) wird anhand eines Multiple-Choice-Fragebogens mit drei Distraktoren pro Vogel überprüft. Der Test von Nahrung und Habitat folgt einer Zuordnungsaufgabe. Jedem Objekt muss die korrekte Merkmalsinformati-on zugeordnet werden. Im Gegensatz zum Test über das Aussehen sind hier alle Vögel und alle Antwortalternativen gleichzeitig aufgelistet und beinhalten keine Distraktoren. Die For-mulierung der Antwortalternativen ist exakt gleich wie auf den Lernkarten. Zur Anwendung kommen drei Fragebögen mit unterschiedlichem Umfang: ein 12er Fragebogen für t1, ein 20er für t2 und t3 und ein 14er für t4.

9.3. Resultate

9.3.1. Überprüfung der Voraussetzungen

Stresswert. Die Interpretierbarkeit von NMDS-Karten ist gewährleistet, wenn die Stresswerte (Gütemaß zur Abschätzung der Abbildungsgüte der Objektrelationen im Vergleich zu den Proximitäten) der NMDS-Lösungen einen möglichst geringen Wert aufweisen¹⁷. In Abbildung 69 sind die Stresswerte der verschiedenen Erhebungszeitpunkte aufgeführt.

Die 12er-Karten aus t1 (\bar{A}_{12-1} , \bar{A}_{12-2}) weisen zwar im Mittel genügend tiefe Stresswerte auf, jedoch streuen die Werte stark. Bei vier individuellen Karten ist der Stresswert, basierend auf früheren Erfahrungen, eher zu hoch. Die zwei 20er-Karten aus t2 (\bar{A}_{20-1} , \bar{A}_{20-2}) und die beiden aus t3 (\bar{A}_{20-3} , \bar{A}_{20-4}) können allesamt als genügend tief angesehen werden. Charakteristisch ist bei den 20er-Karten, dass über die Zeit der mittlere Stresswert und dessen Varianz abnehmen und bei der letzten in Abbildung 69 aufgeführten Erhebung ein sogar sehr tiefes Niveau an-nehmen. Die Tendenz, dass über die Lernzeit und vor allem nach einiger Latenzzeit nach dem Lernen die Konsistenz der Urteile zunimmt und somit der Stresswert der NMDS-Lösungen abnimmt, konnte bereits in anderen Studien nachgewiesen werden (vgl. Kap. 2-5 in dieser Arbeit). Für den Zeitpunkt t4 und die dort berechneten Karten bestehend aus 34 Objekten können keine Stresswerte angegeben werden, da die 34 Objekte nicht gemeinsam, sondern sequentiell mit RobuScal skaliert wurden (vgl. Kap. 9.2).

¹⁷ Die von Gigerenzer (1981) vorgeschlagenen Grenzwerte aus Monte-Carlo-Simulationen setzen bei 20 Objek-ten und zweidimensionaler NMDS-Lösung bspw. einen oberen Wert von 0.28 an (jedoch mit dem MINISSA-Algrithmus).

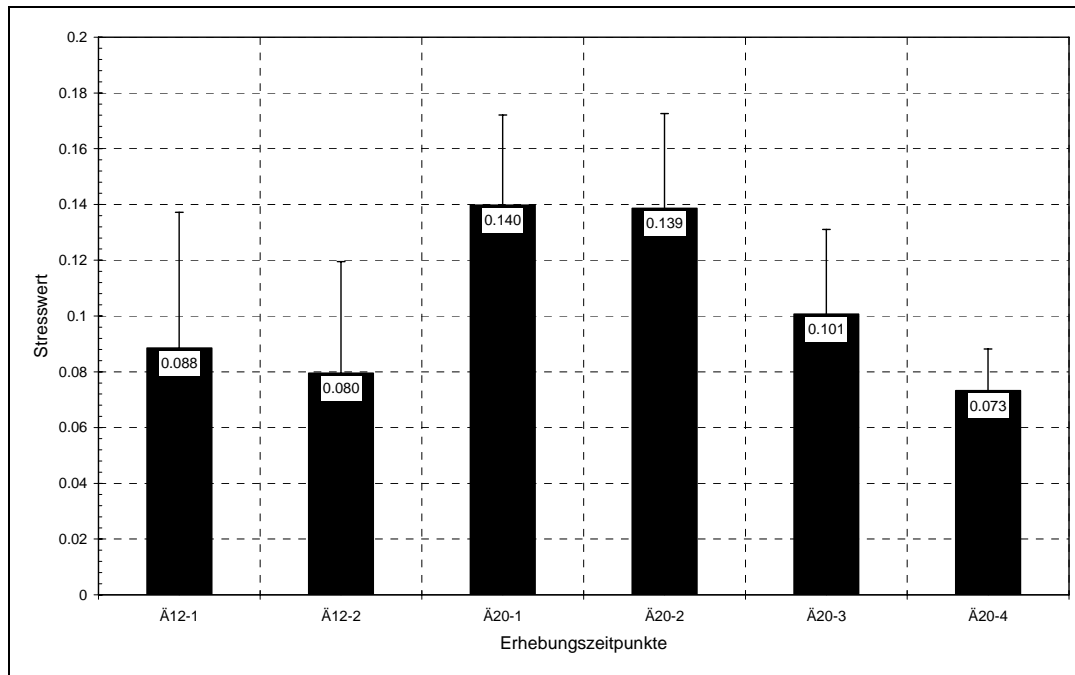


Abbildung 69. Mittelwert und Standardabweichung der Stresswerte der 2-dimensionalen NMDS-Lösungen getrennt nach Erhebungszeitpunkten.

Faktenwissen und Kartenqualität. Im Fragebogen in t1 (F_{12}) mussten, um von genügend gespeichertem Faktenwissen ausgehen zu können, mindestens 75% der Merkmalsfragen korrekt gelöst worden sein, um erstmals Ähnlichkeitsurteile zu 12 Objekten abgeben zu können. Erreichte ein Proband dieses Kriterium nicht, musste er erneut die 12 Objekte repetieren. Die Vpen 2, 3, 4 und 5 durchliefen diese Zusatzrepetition. Um zu gewährleisten, ob sich das memorisierte Sachwissen auch strukturell in den NMDS-Karten widerspiegelt und somit von stabilen und differenziertem Vorwissen ausgegangen werden kann, sollten die Karten der Probanden einen AvgLoss-Wert von 0.3 im Vergleich zur Expertenkarte nicht überschreiten. Fünf Probanden erreichten dieses Kriterium. Eine Person (Vp 4) wies auch nach dreimaligem Lernen und zweifacher Urteilsabgabe immer noch einen hohen AvgLoss-Wert von 0.52 auf. Aufgrund der sehr geringen Anzahl Versuchspersonen wurde diese Person dennoch nicht aus dem weiteren Experimentalverlauf ausgeschlossen.

Der Zeitpunkt t3 hatte zum Zweck, alle sechs Probanden auf ein expertennahes Wissens- und Strukturierungsniveau über die 20 Objekte der Grundstruktur zu bringen, um in t4 von einem stabilen und homogenen Vorwissen ausgehen zu können. Auch hier mussten die Probanden im Fragebogen (F_{20-3}) mindestens 75% korrekte Lösungen erreichen. Ein Proband (Vp 1) erreichte dieses Kriterium nur knapp zur Hälfte (38%), eine weitere (Vp 6) knapp nicht (68%). Im Mittel erreichten die Probanden 81%. Mit einer Ausnahme (wiederum Vp 1), welche den AvgLoss-Wert von 0.41 produzierte, unterschritten alle Probanden das AvgLoss-Kriterium von 0.3 deutlich. Vp 4, welche bei t1 die Kriterien nicht erreichte, produzierte hier mehr als 90% korrekte Lösungen im Fragebogen und unterschritt das AvgLoss-Kriterium von 0.3 mit einem Wert von 0.24 markant. Die gesetzten Kriterien in Fragebögen und Karten wurden insgesamt nicht wie gewünscht erreicht.

Zusammenhang zwischen Merkmalswissen und Kartenqualität. Personen, welche im Fragebogen wenige Punkte erzielen, sollten in ihren Kartenabweichungen vom Expertenmodell ebenfalls höhere Werte generieren als viel wissende Personen. Diese Zusammenhangsvoraussetzung muss gegeben sein, um NMDS-Lösungen als Maß für Wissen überhaupt verwenden zu können. Dass anhand der Abweichungen von Kognitiven Karten zwischen Lerner- und Expertenlösung das Faktenwissen prognostiziert werden kann, wurde bereits in anderen Studien nachgewiesen (vgl. Kapitel 4 in dieser Arbeit). Dort betrug der korrelative Zusammenhang zwischen Punkten in einem Fragebogen und AvgLoss-Werte den signifikanten Wert von -0.49 . In der vorliegenden Studie bietet sich dieser Vergleich zu zwei Zeitpunkten an: Der Zusammenhang zwischen Fragebogen (F_{20-1}) und den AvgLoss-Werten (\bar{A}_{20-1}) der ersten Urteilerhebung in t2 beträgt $r = -0.29$ ($p = 0.29$, einseitig), während in t3 die Produkt-Moment-Korrelation einen viel höheren Wert von $r = -0.77$ ($p < 0.05$, einseitig) aufweist. Dieses Verhältnis ist insofern erstaunlich, als dass in t3 die Personen über etwa ähnlich gutes Wissen im Fragebogen verfügen wie in t2. Ein Hinweis darauf, dass kognitive Strukturierungsleistungen mehr Zeit und aktives Überlegen benötigen als das Lernen von Fakten und sich somit erst nach einiger Latenzzeit Wissensunterschiede auch in der kognitiven Organisation bemerkbar machen? Trotz der eingängigen Logik dieser Interpretation ist sie hier wohl kaum angebracht, denn durch die geringe Anzahl an Versuchspersonen hat ein einzelner Ausreißer bereits einen hohen Einfluss auf das Korrelationsmaß. Vp 1 ist ein solcher Kandidat und steht in t3 sowohl im Fragebogen mit lediglich 38% korrekten Lösungen und einem hohen AvgLoss von 0.41 im Vergleich zu den anderen Personen weit dahinter und dürfte der Grund für die hohe negative Korrelation sein.

Stabilität der NMDS-Lösung bei Erweiterung der Grundstruktur. In Kapitel 5 wurde ein Verfahren zur Reduktion der für die Wissensdiagnostik benötigten Ähnlichkeitsurteile vorgestellt. Eines der Kriterien für die methodische Tauglichkeit war es, dass sich beim Einfügen von neuen Objekten in eine bestehende Struktur die Basisobjekte (Grundstruktur) bzw. deren relationale Ordnung nicht maßgeblich verändern darf. Anders ausgedrückt ging es um die Frage, wie sich die NMDS-Lösung verhält, wenn dieselben Objekte mit denselben Ähnlichkeitsurteilen einmal für sich alleine und demgegenüber mit zusätzlichen Objekten skaliert werden. In der simulativen Studie in Kapitel 5 ergab sich ein Richtwert von AvgLoss = 0.17 bei 15 Objekten der bestehenden Grundstruktur und einer Erweiterung um 10 Objekte. Der Wert sank auf AvgLoss = 0.11 bei einer Grundstruktur von 20 Objekten. Einen ähnlichen Wert erwarten wir auch in dieser Studie mit einem gewissen Spielraum nach oben aufgrund des komplexeren Wissensgebietes und realer Personen im Gegensatz zu berechneten, konsistenten Daten in der erwähnten Untersuchung. Die Daten aus t4 zeigen, dass die 20er-Grundstruktur sich nur marginal verändert, wenn zusätzlich 14 Objekte mitskaliert werden. Der Unterschied zwischen der 20er-Karte skaliert mit 20 Objekten und 20er-Karten skaliert mit 20 plus 14 Objekten beträgt $M = 0.06$ ($SD = 0.03$), womit die Daten sogar unter dem scharfen Kriterium aus der Simulationsstudie liegen. Das Verfahren der sequentiellen methodischen Erweiterung von NMDS-Karten ist stabil.

9.3.2. *Adaptives Lernen vs. Standard-Lernen*

Bevor die Analyse des Vorteils adaptiver Auswahl von Lernmaterialien behandelt wird, folgen zunächst einige allgemeine Resultate zu Lerneffekten der gesamten Stichprobe.

Lernfortschritt im Fragebogen. Erwartet wird ein Wissenszuwachs, welcher sich in steigenden Punktzahlen über die Zeitpunkte ausdrückt. In den zwei Fragebögen in t2 und dem einen in t3 verändern sich die mittleren Punktzahlen ausgehend von $M = 44.5$ ($SD = 3.7$) über $M = 50.2$ ($SD = 4.5$) hin zu $M = 48.7$ Punkten ($SD = 14.3$) bei einer maximal erreichbaren Punktzahl von 60. Generell ist kein klarer Lern-Effekt über alle Zeitpunkte festzustellen (Friedman: $\chi^2(2, 6) = 5.30, p = 0.07$). Obwohl insgesamt zwischen dem ersten und zweiten Zeitpunkt ein Punktezuwachs im Fragebogen festzustellen ist und auch statistisch signifikant wird (Wilcoxon: $Z = -2.20, p < 0.05$, einseitig), ist der Effekt doch eher klein. Wichtig zu bedenken ist hier, dass im F_{20-1} theoretisch lediglich 36 Punkte erzielt hätten werden sollen (das verfügbare Vorwissen der Probanden aus t1 umfasste 12 Objekte mit je drei Merkmalen). Im Mittel erzielten die Probanden also bereits 8 Punkte mehr, ohne aber über zusätzliches Wissen zu verfügen (ein Hinweis darauf, dass die Trennschärfe der Distraktoren in den Fragebögen zu bemängeln sein dürfte). Eine Zunahme von knapp sechs Punkten nach zweimaligem Lernen von je 8 Objekten (24 Merkmalsinformationen) ist sehr wenig, ungeachtet dessen, dass die Auswahl der Objekte bzw. zu lernenden Vögel basierend auf dem Kartenvergleich im Schnitt zwischen rund zwei und sechs mangelhafte Wissens Elemente detektierte (vgl. später die Resultate zur Trefferrate). Die Punktzahl im dritten Fragebogen (t3) bleibt im Vergleich zum zweiten gleich bzw. sinkt sogar ein wenig (eine zufällige Schwankung, die auch durch die hohe Varianz und die kleinen Stichprobengröße zu erklären sein dürfte). Trotz des nochmaligen Repetierens kommt es also nicht zu einer weiteren Verbesserung des Merkmalswissens. Insgesamt können die erwarteten Effekte des Lernfortschrittes nicht gefunden werden.

Konvergenz der NMDS-Lösungen. In zwei früheren Studien (Kapitel 2 in dieser Arbeit) mit unterschiedlichen und verschieden komplexen Inhaltesgebieten zeigte sich eine deutliche konvergierende Tendenz der NMDS-Lösungen in Richtung auf ein Expertenmodell über reines Merkmalslernen. Diesen Effekt finden wir ebenfalls in dieser Studie (Friedman: $\chi^2(3, 6) = 8.60, p < 0.05$; vgl. Abbildung 70), wobei berücksichtigt werden muss, dass neben dem Lernen von Merkmalsinformationen zwischen dem dritten und vierten Erhebungszeitpunkt auch eine Kognitive Karte (das Expertenmodell) als didaktisches Hilfsmittel Verwendung fand. Dort ist auch der einzige deutliche Knick in der Verbesserung der Kognitiven Karten zu sehen, was sich deckt mit subjektiven Aussagen einiger Probanden, dass sie primär die visuelle Information gelernt haben. Die Probanden verstanden es, zusätzlich zu den gesamten vorliegenden Merkmalsattributen die visuell-relationale Information als Strukturierungshilfe zu nutzen. Die Lernsituationen zwischen \check{A}_{20-1} und \check{A}_{20-2} (8 Objekte) und zwischen \check{A}_{20-2} und \check{A}_{20-3} (8 Objekte plus 2-3 Tage Lernpause dazwischen) zeigten jedoch keinen großen bzw. gar keinen Effekt auf die Angleichung der individuellen Karten hin zu einem Normmodell (man beachte die leicht steigende Streuung über die ersten drei Zeitpunkte).

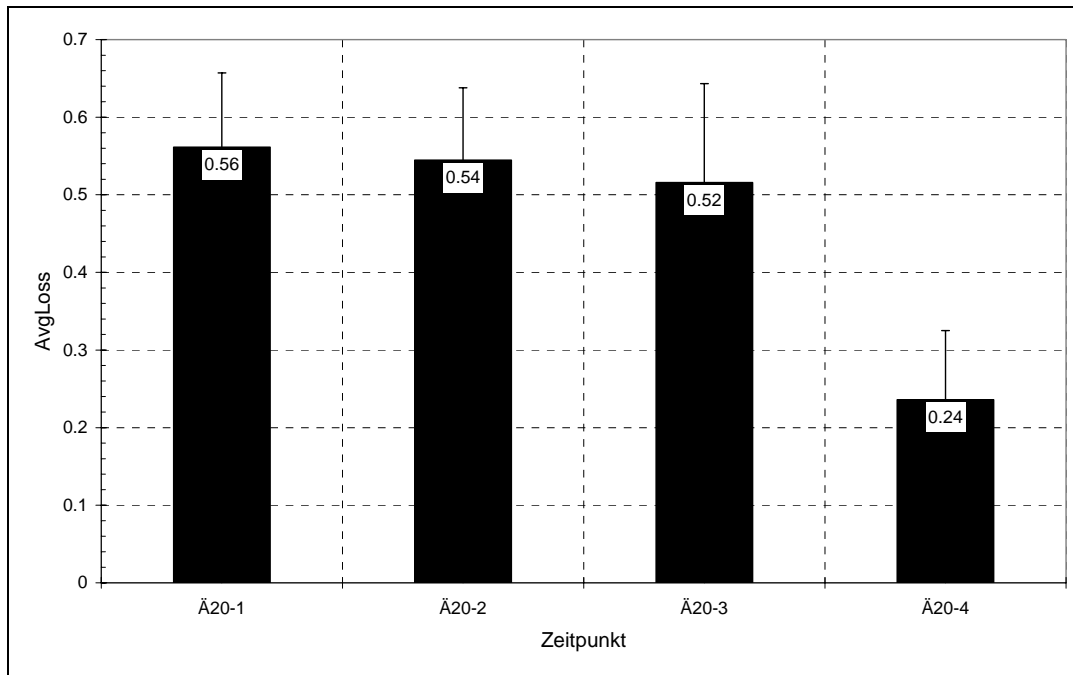


Abbildung 70. AvgLoss-Werte zu unterschiedlichen Zeitpunkten (t2: Ä₂₀₋₁, Ä₂₀₋₂; t3: Ä₂₀₋₃, Ä₂₀₋₄) im Total über beide Stichproben.

Adaptiv vs. Standard. Die eben beschriebene Abnahme der individuellen Kartenabweichungen im AvgLoss detailliert betrachtet beantwortet die Frage, ob die adaptive Auswahl einen Vorteil im Lernfortschritt im Vergleich zu einer Standard-Reihenfolge auf Kartenebene verursacht. In Abbildung 71 sind deshalb dieselben Daten aus Abbildung 70 getrennt nach Experimentalgruppen aufgeführt.

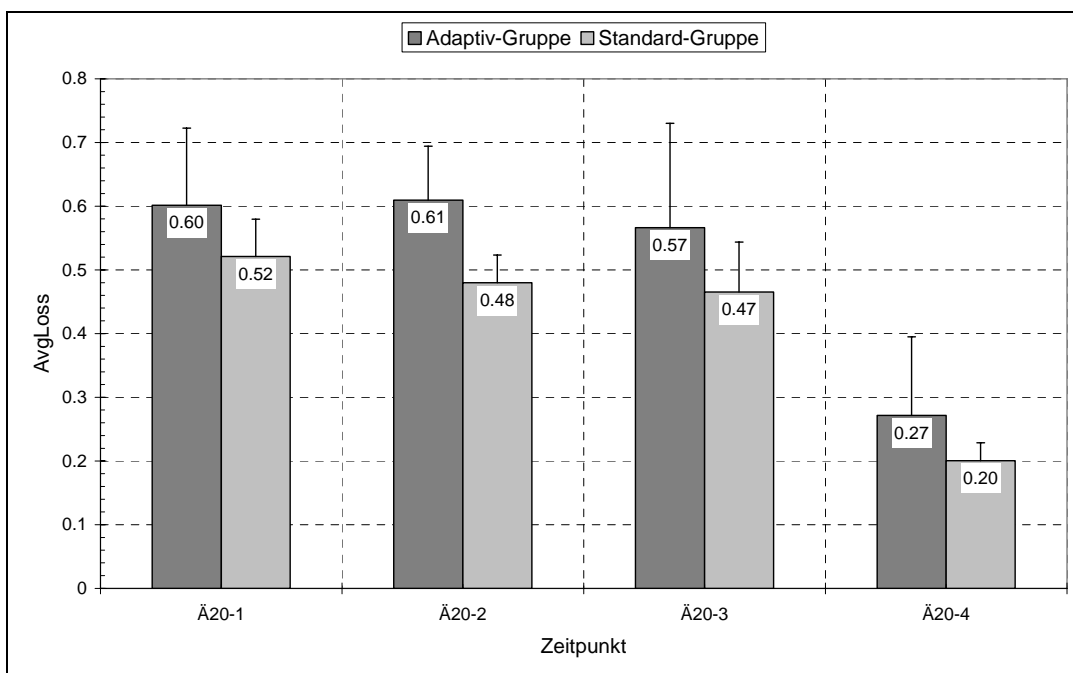


Abbildung 71. AvgLoss-Werte zu unterschiedlichen Zeitpunkten (t2: Ä₂₀₋₁, Ä₂₀₋₂; t3: Ä₂₀₋₃, Ä₂₀₋₄), getrennt nach Experimentalgruppen.

Der Vollständigkeit halber sind in Abbildung 71 alle Abweichungen der individuellen Karten zum Expertenmodell aufgeführt. Ein Einfluss der adaptiven Auswahl der Lerninhalte sollte sich nur zwischen \check{A}_{20-1} und \check{A}_{20-2} und zwischen \check{A}_{20-2} und \check{A}_{20-3} zeigen lassen (zwischen diesen Erhebungszeitpunkten fanden adaptive und standardisierte Lernsitungen statt). Zu sehen ist wenig. Der anfängliche geringe Niveau-Unterschied zwischen den Treatment-Gruppen bleibt über die Zeit konstant bestehen und schwankt zwischen einem AvgLoss von 0.07 und 0.13, wobei die Adaptiv-Gruppe durchgehend schlechter abschneidet als die Standard-Gruppe. Solche Größen müssen als Zufallsschwankungen betrachtet werden, zeigen sich doch in Stabilitäts- und Reliabilitätsstudien zu erwartende intraindividuelle Schwankungswerte im AvgLoss, welche durch den Erhebungszeitpunkt erklärt werden können, von bis zu 0.21 (vgl. Kapitel 5), nicht zu vergessen, die große Ausreißeranfälligkeit bei nur drei Probanden pro Stichprobe. Die deskriptive Analyse zeigt bereits, dass die adaptive Auswahl auf Kartenebene keinen Effekt ergeben hat. Demzufolge wird auf eine nicht-parametrische Analyse verzichtet. Dieser Niveau-Unterschied bleibt des Weiteren auch bei \check{A}_{20-4} in t3 bestehen, obwohl hier beide Gruppen dasselbe Treatment durchlaufen haben, was einzig darauf hinweist, dass in der Standard-Gruppe die drei Personen wohl im Mittel die etwas besseren Strukturierer sind.

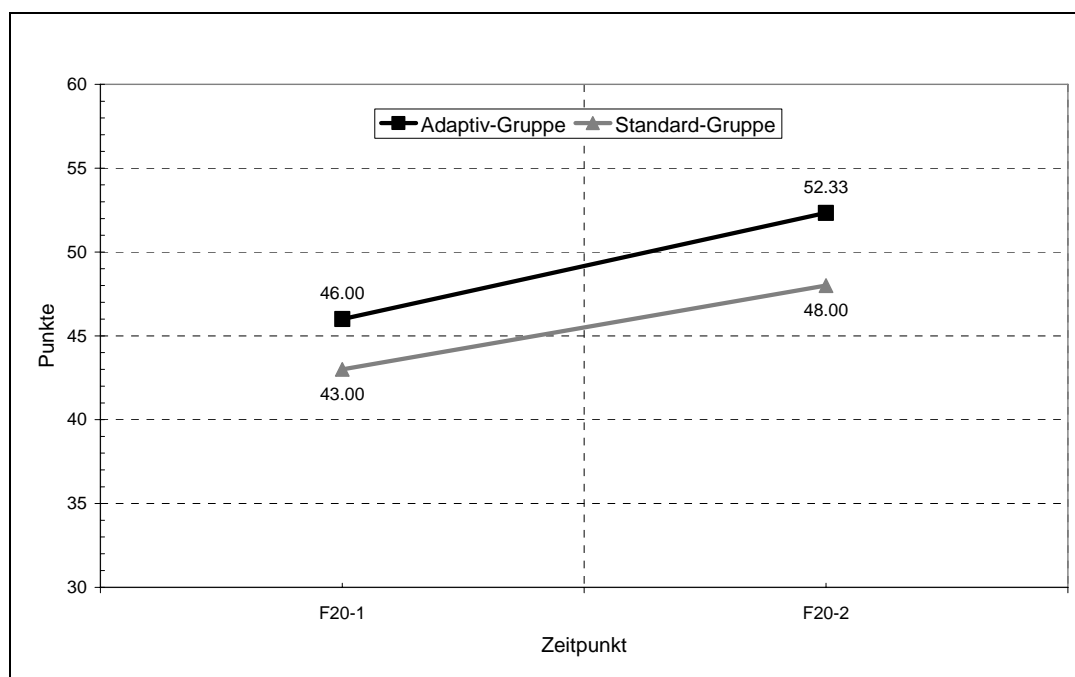


Abbildung 72. Punkte in den beiden Fragebögen in t2 getrennt nach Experimentalgruppe.

Neben dem Effekt auf Kartenebene hätte sich ein Vorteil der Adaptiv-Gruppe auch oder sogar vor allem im Merkmalswissen niederschlagen sollen, gemessen über die Fragebögen F_{20-1} und F_{20-2} in t2 (dazwischen lagen zwei kritische Lernsitungen). Die Adaptiv-Gruppe sollte aufgrund der individuell zusammengestellten Lektionen (im besten Fall ausschließlich) nicht gewusstes Lernmaterial (8 Objekte) bekommen und somit einen größeren Lernerfolg zeigen (vgl. Abbildung 72). Der noch insgesamt betrachtete überzufällige Lernzuwachs im Fragebogen wird nicht mehr signifikant, wenn man die Experimentalgruppen einzeln betrachtet (Wilcoxon: $Z = -1.60$, $p = 0.05$, einseitig, gilt für beide Gruppen), ebenso wenig signifikant ist der

Unterschied zwischen den Gruppen beim ersten Zeitpunkt (Mann-Whitney: $Z = -0.66$, $p = 0.25$). Infolge der adaptiven Auswahl sollte zum zweiten Zeitpunkt die Adaptiv-Gruppe besser abschneiden als die Standard-Gruppe, dies ist jedoch nicht der Fall (Mann-Whitney: $Z = -0.66$, $p = 0.25$).

9.3.3. Prognosequalität der ObjLoss-Werte

Mittelwertsunterschiede bekannt vs. unbekannt. Die adaptive Auswahl von Lernmaterialien funktioniert nur dann, wenn die individuellen ObjLoss-Werte einer Prokrustes-Transformation die Wissensqualität prognostizieren können. Objekte und deren Abweichungen zu einem Expertenmodell, welche der urteilenden Person gänzlich unbekannt sind bzw. noch nie gelernt wurden, sollten systematisch höhere Werte aufweisen als ObjLoss-Werte von bereits bekannten Objekten¹⁸. Dann kann eine automatisierte adaptive Auswahl von noch zu lernenden Objekten erfolgreich sein. Dass dies der Fall ist, konnte bereits in einer experimentellen Lernstudie und einer simulativen Untersuchung gezeigt werden (vgl. Kapitel 4). Tauglich ist dieses Verfahren aber nach diesen Resultaten nur dann, wenn die Karten mindestens 16 Objekte beinhalten und von einem korrekten Vorwissen von mindestens 60% ausgegangen werden kann.

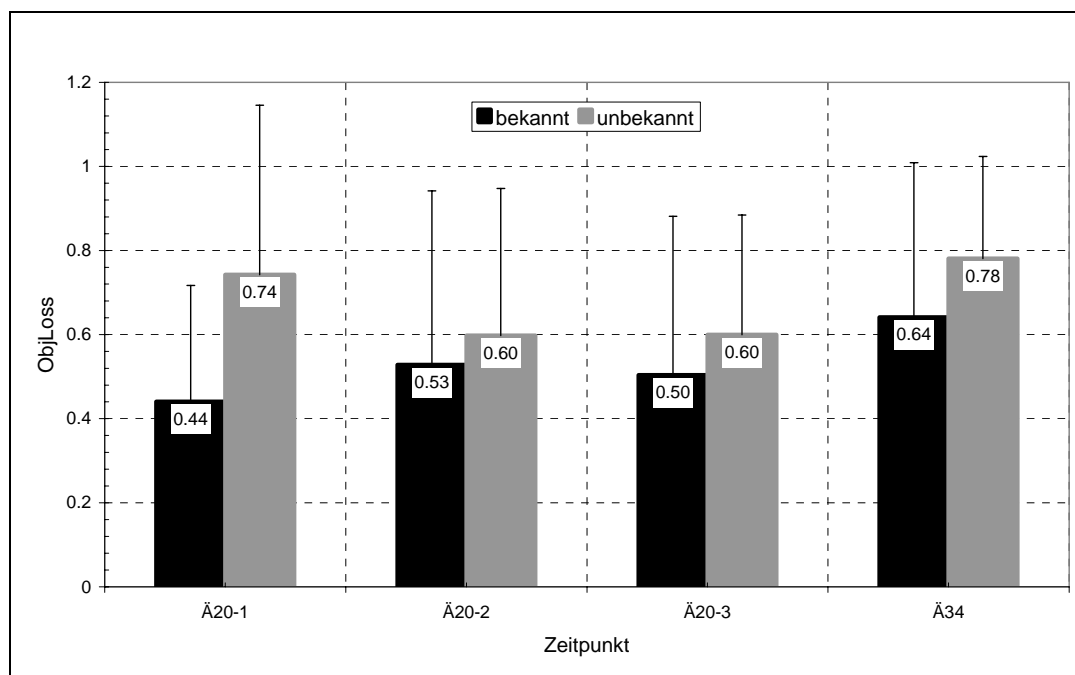


Abbildung 73. Objektabweichungen (ObjLoss) der bekannten und unbekannten Objekte zu verschiedenen Erhebungszeitpunkten.

¹⁸ Genauso wie in den Studien in Kapitel 4 wird bekannt und unbekannt darüber operationalisiert, ob die Merkmalsinformation zu diesem Objekt theoretisch bekannt sein kann, d.h. in einer der Lernsitungen mindestens einmal vorgekommen ist (und nicht, ob das Objekt tatsächlich auch gewusst ist; aufgrund des Experimentalsettings ist diese Klassifikation nicht möglich). Unbekannt bedeutet somit, dass das Lernmaterial zu diesem Objekt noch nie gelernt werden konnte.

In der vorliegenden Studie kann die gerichtete Hypothese bei vier Erhebungen untersucht werden (Abbildung 73). Im Mittel weisen die bekannten Objekte über alle Zeitpunkte hinweg wie erwartet die geringeren ObjLoss-Werte auf. Jedoch werden nur zwei Differenzen statistisch signifikant und die Abweichungswerte der bekannten Objekte bewegen sich ganz allgemein auf einem sehr hohen Niveau: \bar{A}_{20-1} (60% bekannt; $t(76) = -4.53$, $p < 0.001$, einseitig) und \bar{A}_{34} (50% bekannt; $t(82) = -2.05$, $p < 0.05$, einseitig)¹⁹. Bei beiden Erhebungen, \bar{A}_{20-2} (76% bekannt; $t(118) = -0.82$, $p = 0.21$, einseitig) und \bar{A}_{20-3} (88% bekannt; $t(118) = -0.94$, $p = 0.17$, einseitig) kann kein Effekt festgestellt werden.

Trefferrate. Im Hinblick auf die praktische Brauchbarkeit des Verfahrens stellt sich für ein Intelligentes Tutorielles System die Aufgabe, möglichst viele unbekannte Objekte zu entdecken und demgegenüber möglichst wenig bereits bekannte Objekte fälschlicherweise als unbekannt zu klassifizieren. In Analogie der Studien in Kapitel 4 wird auch hier keine Sensitivität von 100% als Kriterium der Tauglichkeit verlangt (unbekannte Objekte, welche verpasst wurden, können auch in einem nächsten Diagnosezyklus noch detektiert werden). Bei einer tolerierten Falsch-Alarm-Rate von 25% (und einer Trefferrate von 75%) zeigt sich in Abbildung 74, dass der proportionale Anteil an falsch klassifizierten Objekten nur beim ersten Zeitpunkt einigermaßen valide ist. Bei einem Vorwissen von 60% (12 von 20 Objekten bekannt) ordnet das Verfahren mehr als einen Drittel der bekannten Objekte (4 von 12) fälschlicherweise der unbekannten Klasse zu. Im Experiment in Kapitel 4 wurde solch ein Wert nie erreicht. Die fälschliche Zuweisung steigert sich auf einen beträchtlichen Wert von etwas unter 60% Falsch-Detektion bei der zweiten und dritten Erhebung. Diese markante Zunahme der Falsch-Alarm-Raten in der zweiten und dritten Erhebung könnte ein Hinweis darauf sein, dass die Klassifikation von bekannt und unbekannt nicht ganz adäquat ist aufgrund der Einteilung eines Objektes in die Kategorie „bekannt“, sobald das Wissen einmal gelernt werden konnte (Die Vpen sagten in mündlichen Nachbefragungen auch aus, dass die Lernzeit sehr kurz war und sie die gelernten Objekte subjektiv nicht alle als gewusst empfanden). Vor der ersten Erhebung haben die Probanden die Merkmalsinformation mindestens drei und maximal fünf Mal gelernt, womit die Wahrscheinlichkeit, dass ein bekanntes Objekt auch wirklich gewusst wird, erheblich höher ist als bei der zweiten und dritten Erhebung. Die Falsch-Alarm-Rate der bekannten Objekte sinkt dann bei der letzten Erhebung auf einen Wert von gut 40%. Dieser bessere Wert in der letzten Erhebung lässt sich mit der durch die größere Anzahl an Objekten (insgesamt 34) erreichten robusteren Positionierung der Objekte in den NMDS-Lösungen erklären. Vergleicht man diese Werte mit dem ähnlich ausgerichteten Experiment in Kapitel 4, erkennt man insgesamt eine hohe Übereinstimmung und sogar leicht optimistischer stimmende Resultate²⁰. Wenn die Sensitivität des Verfahrens auch nicht optimal ist, zeigt sich zumindest eine stabile und replizierbare Treffergüte.

¹⁹ In die Auswertung von \bar{A}_{34} sind lediglich die ObjLoss-Werte der 14 neuen Objekte eingeflossen.

²⁰ In Kapitel 6 wurde ein ähnliches, jedoch nicht ganz ineinander überführbares Vorgehen zur Bestimmung der Trefferrate herangezogen, allerdings mit dem Unterschied, dass dort unter Zuhilfenahme von Kohonen-Netzen das Lernmaterial auf Merkmalsebene zusammengestellt wurde. Die Sensitivität, falsch oder nicht gewusste Merkmale zu treffen, betrug dort zwischen 33% und 45%. Das System detektierte nicht ganz die Hälfte des

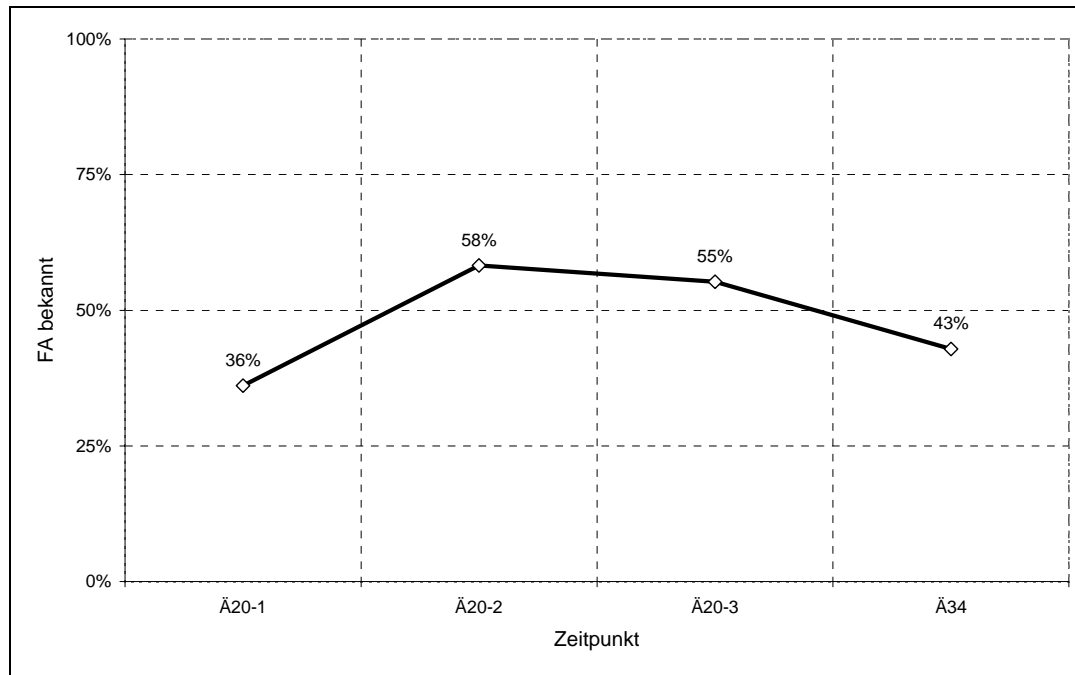


Abbildung 74. Falsch-Alarm-Raten der bekannten Objekte bei einer Sensitivität von 75% für die unbekannten Objekte.

9.4. Diskussion

Zusammenfassung der Resultate. Insgesamt zeigt sich auf den ersten Blick ein sehr heterogenes, undeutliches Bild und die gerichteten Vorhersagen und Erwartungen in das System konnten nur bedingt bestätigt werden. Wegen der einigermaßen großen Anzahl an Daten und Ergebnissen des komplex strukturierten Experimentes wird nachfolgend als Diskussionsgrundlage erst eine prägnante Gegenüberstellung der wichtigsten positiven und negativen Resultate in Bezug auf die Vorhersagen gezeigt (Tabelle 5).

Tabelle 5. Zusammenfassung der Hauptresultate bezogen auf die Bestätigung der Vorhersagen.

positiv	negativ
korrelativer Zusammenhang Fragebogen und NMDS-Karte	kein Lerneffekt im Fragebogen
mittlere Trefferraten von mangelhaftem Wissen	kein Lerneffekt auf Kartenebene
hohe methodische Stabilität der NMDS bei sequentieller Einpassung neuer Elemente	kein Effekt des adaptiven Treatments

Beginnen wir mit den bestätigten Befunden. Anhand der gefundenen negativen Korrelationen zwischen Punktzahlen im Fragebogen und der Abweichungen der individuellen Karten vom Expertenmodell kann wie bereits in anderen Studien (vgl. Egloff, 2004 und Kapitel 2, 4 und 5

mangelhaften Wissens. Da der Anspruch an die Treffergenauigkeit von Merkmalen einiges höher ist als auf Objektebene Mängel zu entdecken, können diese Resultate hier ebenfalls als vergleichbar gut bezeichnet werden.

in dieser Arbeit) eine stabile und reliable Kovarianz der Wissensmessung über zwei verschiedene Diagnoseformate ein weiteres Mal bestätigt werden.

Für die praktische Brauchbarkeit dieses Diagnoseverfahrens ist es darüber hinaus auch notwendig, dass das System auf der Beschreibungsebene „Objekt“ diejenigen detektieren kann, welche von einer Person nicht oder falsch gewusst werden. Die Trefferraten der ObjLoss-Werte bei einer Sensitivität von 75% sind nicht perfekt, aber akzeptabel und vergleichbar hoch wie in einer bereits durchgeführten experimentellen Studie (Kapitel 5 in dieser Arbeit). Außerdem kann prognostiziert werden, dass über eine bessere Operationalisierung der Kategorien „bekannt“ und „unbekannt“ die Treffergenauigkeit noch ansteigt. Das System muss auch nicht perfekt sein. Ein ITS oder ATS zeichnet sich durch seine zyklische Struktur aus, d.h., dass immer wieder ein Diagnoselauf absolviert wird und somit mangelhaft gewusste Objekte später erneut „die Chance kriegen“, entdeckt und für die kommende Lerneinheit empfohlen zu werden.

Des Weiteren scheint sich die sequentielle Einpassung von Objekten in eine bestehende NMDS-Struktur als tauglich zu erweisen. Der Einfluss von zusätzlich zu den Objekten der Grundstruktur skalierten Objekten auf die relationalen Verhältnisse war in diesem Experiment sogar noch kleiner als bei simulierten Analysen in einem anderen und einfacher strukturierten Inhaltsgebiet (Egloff, 2004). Setzt man ein stabil verankertes und für das Wissensgebiet repräsentatives Grund- oder Überblickswissen und somit auch eine Grundgerüst für weiteres assimilatives Lernen voraus, kann im weiteren Lernverlauf auf die Beurteilung bekannter Objekte verzichtet werden. Die Diagnostik greift trotzdem valide, auch wenn nur noch Ähnlichkeitsbeurteilungen neu gelernter Elemente zu bekannten befragt werden. Dies ist gleichbedeutend mit einem erheblichen Effizienzgewinn im Diagnostikaufwand, der sich trotz des immer noch beträchtlichen Zeitbedarfs bei vielen Objekten so zumindest überhaupt erst als praktikabel und für die Nutzer zumutbar erweist²¹.

Die hauptsächlich negativen Befunde überwiegen die positiven in ihrer Schwere die Ziele dieses Experimentes betreffend leider erheblich. Als grundsätzliche Bedingung eines Experimentes, in welchen Personen wiederholt Wissensdaten lernen sollen, gilt eine merkliche Verbesserung deren Wissensstandes. In den Multiple-Choice-Fragebögen ist das Fazit trotz der leichten Zunahme der Punktzahlen im zweiten Fragebogen ernüchternd, sinkt der Wert im dritten dann trotz dazwischen liegender Lernsituationen sogar wieder ab. Diese sind so gering, dass unter Berücksichtigung der geringen Stichprobengröße und der vermuteten schlechten Trennschärfe der Frage-Items leider diese pessimistische Schlussfolgerung gezogen werden muss.

²¹ Die Tauglichkeit des sequentiellen Einpassens führt dann sogar zu einem potentiellen Diagnostikkonzept, welches in Anlehnung an das „gleitende Testfenster“ von Leutner (1993) die zu beurteilenden Objekte einzeln und in den Lernprozess integriert diagnostizieren lässt. Ein Lerner beurteilt also je ein neues Objekt mit denjenigen der GS. Bei einer GS von 20 Objekten sind das 20 Urteile, für welche nicht mehr als 2 Minuten benötigt werden. Eine klare Trennung zwischen Lern- und Diagnoseeinheiten ist nicht mehr notwendig.

Auf Kartenebene sind die Resultate nur leicht besser. Die mittlere individuelle Kartenabweichung zur Norm zu Beginn der zweiten Sitzung mit einem AvgLoss rund 0.56 stimmt überein mit einem mittleren Faktenwissen von 60% (vgl. Kapitel 4). Eine kontinuierliche Abnahme zwischen erster und dritter Diagnose in t2 und t3 ist zwar vorhanden, jedoch muss auch hier wieder die geringe Stichprobengröße vorsichtig stimmen, die Resultate nicht über zu bewerten. Erklärungsbedarf verlangt die starke Abnahme bei der letzten Diagnose in t3. Bis auf eine Versuchsperson erreichten alle das gesetzte Kriterium und wichen von der Normkarte mit einem AvgLoss-Wert von deutlich weniger als 0.3 ab (Dies schafften in anderen Experimenten mit ähnlich komplexem Wissensgebiet nur wenige Personen.). Dieser geringe Wert ist umso erstaunlicher, als dass lediglich eine Lernsitzung von rund 20 Minuten zwischen den beiden Ähnlichkeitserhebungen in t3 dazwischen lag. Wichtig zu bedenken ist nun aber, dass während dieser Lernsitzung die Personen die Expertenkarte als visuelle Lernhilfe zu Verfügung hatten. Dies schien für die Probanden eine Art Erlösung zu sein, sich die schwierigen Merkmalssätze zu merken. In der unstrukturierten und freien Nachbefragung merkten drei der sechs Vpen unaufgefordert an, dass sie sich nicht mehr auf das eigentliche Lernmaterial konzentrierten, sondern sich die Kognitive Karte und die relationalen Verhältnisse so gut wie möglich einprägten. Es scheint ihnen gelungen zu sein. Anscheinend sind wir sehr gut in der Lage, aufgrund geometrischen Wissens zueinander passende und objektiv richtige Ähnlichkeitswerte zu bilden.

Die eigentliche Ernüchterung dieses Experimentes besteht aber darin, dass kein Effekt des adaptiven Treatments im Vergleich zu einer Standardreihenfolge im Lernprozess gefunden werden konnte. Weder im Fragebogen noch in den Kognitiven Karten gibt es Unterschiede zwischen den beiden Experimental-Gruppen. In einer anderen Studie (vgl. Kapitel 6 in dieser Arbeit) wurde bereits ein erster Versuch unternommen, mit dem Ziel der Replikation der Resultate aus einer Simulationsstudie, den Vorteil adaptiven Auswählens nachfolgender Lektionen empirisch zu belegen. Darin war die Anforderung an das Intelligente Tutorielle System sogar noch höher, wurden die Lektionen ja nicht nur aufgrund der Objektabweichungen in NMDS-Karten zusammengestellt, sondern unter Zuhilfenahme von Kohonen-Netzen die Sensitivität des Verfahrens auf Merkmalsebene zu bringen. Ein erhöhter Lernfortschritt blieb jedoch aus und die Interpretation aufgrund der Befunde lag nahe, dass die Lerner schlichtweg überfordert waren, jeweils nur einzelne Merkmale in Lektionen ungeordnet vorgelegt zu bekommen und sogleich in das bestehende Wissen zu integrieren. Aus diesem Grunde erhielten in der vorliegenden Studie die Personen die vermeintlich „leichtere“ Strukturierungsaufgabe, jeweils immer Objekte als Ganzes zu lernen, ungeachtet dessen, dass eine Person eventuell bereits Bekanntes erneut zur Verfügung gestellt bekam. Aber auch dieser Wunsch blieb leider unerfüllt. Erneut kann vermutet werden, dass die Lerner überfordert waren mit dem Experimentalablauf. Die Lernzeit war zu kurz, das Vorwissen noch nicht wirklich stabil verankert und die Anzahl kritischer Treatments und somit die Datenbasis zu klein, überhaupt Effekte zu produzieren und sie dann statistisch auch noch zu entdecken.

Mögliche Fehlerquellen. Im Folgenden sollen verschiedene Hypothesen aufgestellt und soweit möglich bereits beantwortet bzw. ausgeschlossen werden, wo die Ursachen für die schlechten Resultate in diesem Experiment gelegen haben könnten (einige wurden ja bereits

im Verlauf der Arbeit angesprochen). Auch hier soll dies der besseren Übersichtlichkeit halber ausgehend von einer tabellarischen Auflistung dargestellt werden (Tabelle 6).

Tabelle 6. Liste möglicher Ursachen für die schlechten Resultate im Experiment.

Mögliche Fehlerquellen	als Ursache ausgeschlossen?
sehr hohe oder sehr tiefe Stress-Werte	ja
schlechte Normkarten	ja
zu geringe Trefferrate	ja
schlechte Fragebogenitems	nein
zu komplexes Experiment	nein
zu geringe Stichprobengrösse	nein

Die naheliegenste Vermutung bei einem Experiment, welches von seiner Ausrichtung her stark auf methodische Aspekte ausgerichtet ist, wieso begründete Hypothesen nicht eingetroffen sind, liegt natürlich in der Tauglichkeit der Methode(n) selbst. Eine erste Schwierigkeit könnte darin liegen, dass die NMDS-Prozedur die Proximitäten nur mit ungenügender mathematischer Qualität in das geometrische Raummodell gegebener Dimensionalität überführen konnte und somit die Objektpositionierungen nur ein Artefakt darstellen. Hinweise darauf hätten sehr große oder gegen Null gehende Stress-Werte der NMDS-Lösungen gegeben. Letzteres hätte auf degenerierte NMDS-Karten hingewiesen. Dies ist aber nicht der Fall.

Genauso wenig kann die Qualität der Normkarten in Frage gestellt werden. Die verwendeten Normkarten sind Resultate von Befragungen von ausgewiesenen Fachexperten und haben sich bereits in anderen Studien bewährt. Einziger Kritikpunkt könnte sein, dass die 34er-Normkarte, welche in t4 verwendet wurde, kein valides Vergleichsmodell sein könnte aufgrund dessen, dass die Experten nicht in klassischer Weise sequentiell paarweise Ähnlichkeiten beurteilten, sondern in eine bestehende, graphisch dargestellte NMDS-Lösung von Hand Punkte eintrugen. Dieses Vorgehen wurde bislang noch nie auf seine Qualität hin geprüft und bietet sich als mögliche Fehlerquelle demzufolge an. Dagegen spricht aber, dass die Trefferraten der ObjLoss-Werte vergleichbar gut oder sogar noch etwas besser sind wie in den 20er-Karten. Wären das Vergleichsmodell schlecht und die Objektrelationen nicht adäquat, müssten die Falsch-Alarm-Raten der gewussten Objekte gegen die Zufallserwartung von 75% gehen. Die Trefferraten sind aber, vergleicht man sie mit einer früheren empirischen Studie (Kapitel 4), ähnlich hoch. Dass die adaptiven Lerner in t2 praktisch keinen feststellbaren Lerngewinn erzielt haben (genauso wenig wie die andere Stichprobe), kann somit auch nicht mit mangelnder Treffergenauigkeit erklärt werden. Aufgrund dessen hätte viel mehr Punktezuwachs im Merkmalsfragebogen erreicht werden können. Die Trefferraten sind nicht sehr gut und für den Anspruch der Anwendung solch eines Verfahrens in naturwissenschaftlichen

Fachgebieten wie der Medizin bspw. gänzlich ungeeignet, weil das Verfahren zu viele Fehler produziert. Für den Zweck der Wissensdiagnostik mit Messwiederholung reichen sie aber aus.

Einzig die Fragebogenkonstruktion kann methodisch betrachtet in Frage gestellt. Die Feststellung, dass die Probanden im ersten Fragebogen in t2 plötzlich mehr Punkte machen, als rein über die gewussten Inhalte hätte möglich sein dürfen, deutet darauf hin, dass die Trennschärfe der einzelnen Items bzw. der Distraktoren mangelhaft war. Vor allem die Zuordnung der Vogel-Abbildungen konnte über das Ausschlußverfahren sehr einfach gelöst werden (in allen Fragebögen waren die Distraktoren real existierende und den Lernern bekannte Merkmalsattribute). Andererseits erzielten die Lerner insgesamt nicht viele Punkte und mussten dies auch nicht (das gesetzte Kriterium von 75% korrekt war viel zu tief gewählt) trotz mehrfachen Repetierens, was zusätzlich zu Aussagen der Probanden, dass die teils sehr ähnlichen Texte nicht auseinander gehalten werden konnten, passt und auf die Schwierigkeit des Inhaltsgebietes hinweist.

Insgesamt betrachtet war das Ziel, mit einem einzigen Experiment all die zahlreichen Fragestellungen und Hypothesen zu bestätigen, ein gewagtes Unterfangen und verlangte den Teilnehmern der Studie eine hohe Leistung ab. Alle Versuchspersonen waren Laien in Ornithologie und beschäftigten sich erstmals in dieser Tiefe mit dem Stoff. Dass dann über eine einzelne Lernsitzung bereits stabiles Vorwissen hätte simuliert werden sollen, muss im Nachhinein stark bezweifelt werden. Ebenso war die Erwartung zu optimistisch, dass über eine zweimalige adaptive Auswahl der Lektionen und einer Lernzeit von insgesamt 30 Minuten aufgrund der Unvertrautheit mit diesem Stoffgebiet bereits ein Adaptivitäts-Effekt auftritt. Die Ursache liegt aber logisch betrachtet dennoch nicht darin, dass die adaptive Auswahl aufgrund mangelnder Treffergenauigkeit der Kartenvergleiche oder grundsätzlich nicht funktionieren kann, sondern einzig darin, dass die Lerner den vorgelegten Stoff nicht zu integrieren verstanden. Einige der Probanden sprachen auch von Konfusion und subjektivem Gefühl des Nicht-Verstehens während des gesamten Experimentes. Weiter zur Komplexität des Experimentes beigetragen hat auch der große Aufwand. Vier Sitzungen à mind. 90 Minuten mit didaktisch nicht optimal aufbereitetem Lernmaterial und wiederholtem Abgeben von überaus zahlreichen Ähnlichkeitsbeurteilungen verteilt über gut zwei Wochen war für die Freiwilligen nicht unbedingt eine Motivationsspritze und sehr ermüdend (Ausnahmslos alle Probanden empfanden das Inhaltsgebiet Ornithologie auch nicht als spannend und das Experiment als großen und frustrierenden Aufwand.).

Wohl der wichtigste Faktor, der aber zu diesen eher unbefriedigenden Resultaten geführt hat, ist wohl die extrem geringe Stichprobengröße. Leider war es nicht möglich, im geplanten Zeitraum mehr freiwillige Versuchspersonen zu finden. Auch wenn es für diesen Gruppenumfang geeignete inferenzstatistische Verfahren gibt und für die Analysen der ObjLoss-Werte sogar parametrische Verfahren verwendet werden konnten, ist die Anfälligkeit für Ausreißer natürlich enorm groß. Im Resultateteil wurde bereits darauf hingewiesen, dass die großen und eigentlich unerwarteten Differenzen der Korrelationswerte zwischen Fragebogen und AvgLoss dadurch verursacht wurden, dass eine einzelne Person in ihrer Leistung hinter den restlichen fünf Personen zurücklag und so die Höhe der Korrelation in die Höhe schnellte. Dasselbe gilt auch insgesamt. Eine einzelne Person kann gegebenenfalls vorhandene Effekte

durch ihr Verhalten vollständig überdecken, aber auch nicht vorhandene Effekte überhaupt als solche erscheinen lassen.

All diese Erkenntnisse können, leicht umgedeutet, als Ausblick dieser Studie betrachtet werden. Leider kann auf inhaltlicher Ebene nicht viel Neues über Nützlichkeit, Performanz und Anwendungsempfehlungen der Nonmetrischen Multidimensionalen Skalierung und Prokrustes-Transformation im Zusammenhang mit adaptiven Lernempfehlungen in Intelligenten, Adaptiven Tutoriellen Systemen oder assimilativen Lernens gesagt werden. Diesbezüglich verweist dieses Kapitel auf die Untersuchungen und deren Diskussionssektionen, auf welchen diese Studie hier beruht (vor allem Kapitel 2, 4, 5 und 6). Für einen Replikationsversuch, welcher die noch nicht abschließend beantworteten Fragestellungen ins Zentrum stellen möchte, soll diese Erkenntnisse aber als Richtwerte verwenden, ein solches Lernexperiment mit großer Sorgfalt zu konstruieren und nicht aufgrund methodischer Versäumnisse fälschlicherweise wichtige und für die zukünftige Bildungsforschung zu innovativen und automatisierten intelligenten Lernsystemen relevante Effekte zu verpassen.

9.5. Literatur

- Albert, D. & Lukas, J. (Hrsg.) (1999). *Knowledge spaces. Theories, empirical research and applications*. Mahwah: Erlbaum.
- Ausubel, D. P. (2000). *The acquisition and retention of knowledge*. Dordrecht: Kluwer Academic Publishers.
- Bezzel, E. (1993). *Kompendium der Vögel Mitteleuropas. Passeres – Singvögel*. Wiesbaden: AULA-Verlag GmbH.
- Charney, D., Reder, L. & Kusbit, G. W. (1990). Goal setting and procedure selection in acquiring computer skills. A comparison of tutorials, problem solving and learner exploration. *Cognition and Instruction*, 7, 323-342.
- Crowder, N. A. (1959). Automating tutoring by means of intrinsic programming. In E. Galanter (Ed.), *Automatic teaching. The state of the art*. New York: John Wiley and Sons, Inc. Pp. 109-116.
- Egli, S., Schlatter, K., Streule, R. & Läge, D. (2006). A structure based expert model of the ICD-10 mental disorders. *Psychopathology* 39 (1), 1-9.
- Egloff, S. (2004). *Effizienz in der ähnlichkeitsbasierten Wissensdiagnostik*. Unveröffentlichte Lizentiatsarbeit, Philosophische Fakultät, Universität Zürich. Zürich.
- Gigerenzer, D. (1981). *Messung und Modellbildung in der Psychologie*. München: Reinhardt.
- Goldstein, I. P. (1982). The genetic graph: A representation for the evolution of procedural knowledge. In D. Sleeman & J. S. Brown (Eds.), *Intelligent tutorial systems* (pp. 51-78). New York: Academic Press, Inc.
- Gower, J. C. & Dijksterhuis, G. B. (2004). *Procrustes problems*. Oxford: Oxford University Press.
- Hudec, K. (1992). *Vögel beobachten und bestimmen*. Hanau/Main: Werner Dausien.
- Klauer, K. J. (1989). Allgemeine oder bereichsspezifische Transfereffekte eines Denktrainings. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 21, 185-200.
- Läge, D. (2001). *Ähnlichkeitsbasierte Diagnostik von Sachwissen*. Habilitationsschrift an der Philosophischen Fakultät der Universität Zürich.
- Läge, D., Daub, S., Bosia, L., Jäger, C. & Ryf, S. (2005). *Die Behandlung ausreißer-behafteter Datensätze in der Nonmetrischen Multidimensionalen Skalierung – Relevanz, Problemanalyse und Lösungsvorschlag*. AKZ-Forschungsbericht Nr. 21. Zürich: Angewandte Kognitionspsychologie.
- Leutner, D. (1992a). *Adaptive Lernsysteme. Instruktionspsychologische Grundlagen und experimentelle Analysen*. Weinheim: Psychologie Verlags Union.

- Leutner, D. (1992b). Das Testlängendilemma in der lernprozess-begleitenden Wissensdiagnostik. *Zeitschrift für Pädagogische Psychologie*, 24, 232-248.
- Leutner, D. (1993). Das gleitende Testfenster als Lösung des Testlängendilemmas: Eine Robustheitsstudie. *Zeitschrift für Pädagogische Psychologie*, 7, 33-45.
- Leutner, D. (1995). Adaptivität und Adaptierbarkeit multimedialer Lehr- und Informationssysteme. In L. J. Ising & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia* (S. 140-149). Weinheim: Psychologie Verlags Union.
- Marx, W., & Hejj, A. (1989). *Subjektive Strukturen*. Göttingen: Hogrefe.
- Marx, W., & Läge, D. (1995). *Der ideologische Ring*. Göttingen: Hogrefe.
- Piaget, J. (1976). *Die Äquilibration der kognitiven Strukturen*. Stuttgart: Klett.
- Pressey, S. L. (1926). A simple apparatus which gives tests and scores and teaches. In (reprint) A. A. Lumsdaine & R. Glaser (1960), *Teaching machines and programmed learning* (pp. 35-41). Washington D.C.: National Education Association of the United States.
- Skinner, B. F. (1958). Teaching machines. In (reprint) A. A. Lumsdaine & R. Glaser (1960), *Teaching machines and programmed learning* (pp. 137-158). Washington D.C.: National Education Association of the United States.
- Smith, E. E., Shoben, E. J. & Rips, L. J. (1974). Structure and process in semantic memory. A featural model for semantic decisions. *Psychological Review*, 81, 214-241.
- Strohschneider, S. (1990). *Wissenserwerb und Handlungsregulation*. Wiesbaden: Deutscher Universitätsverlag.
- United Soft Media (2001). *Der grosse Kosmos-Vogelatlas 3.0* [CD-Rom]. München: Navigo.
- Van Merriënboer, J. J. G., Van Den Berg, K. G. & Maaswinkel, D. M. (1989). Some experiences with two intelligent tutoring systems for teaching computer programming: PROUST and LISP-tutor. J. M. Pieters (Ed.), *Intelligent tutorial systems and instruction* (pp. 7-44). Enschede: OTG Onderwijsleerprocessen.

Danksagung

Der Dank im Zusammenhang mit solch einer Schrift, welche während dreier Jahre intensiver projektorientierter Teamarbeit im Rahmen eines umfassenden Forschungsprogramms entstanden ist, gebührt selbstverständlich auch all jenen, welche hier nicht genannt sind.

Ausdrücklich meinen allerherzlichsten Dank aussprechen möchte ich ...

- ... Prof. Dr. Damian Läge, für die hervorragende, aufopfernde und geduldige Ausbildungs- und Betreuungsarbeit während meiner gesamten Lizentiats- und Doktorandenzeit, die unzähligen konstruktiven und anregenden Diskussionen, ohne die diese Arbeit nicht entstanden wäre.
- ... Katharina Schlatter, Iris Rüfenacht, Stefan Ryf, Samy Egli, Stephan Egloff, Andy Sicker und Christian Fichter welche zusammen mit mir die lebendige Forschergruppe geprägt haben und mich in vielerlei Hinsicht während der Promotionszeit Unterstützung geboten haben.
- ... meinen Eltern für das Vertrauen, das sie einem Jungwissenschaftler entgegengebracht haben und vor allem Nicole Sigrist für all die motivierenden und aufbauenden Worte und ihrer stetigen, jahrelangen Unterstützung.

Curriculum Vitae

Personalien

Name	Roland Streule
Geschäftsadresse	Binzmühlestrasse 14 / Box 28, 8050 Zürich
Telefon G.	+41 (0)44 635 74 41
E-Mail	r.streule@psychologie.uzh.ch
URL	http://www.pto.uzh.ch
Geboren	2. Juli 1976
Heimatort	Appenzell (AI)
Zivilstand	ledig



Ausbildung

2002-2004	Doktorand an der Universität Zürich, Psychologisches Institut, Angewandte Kognitionspsychologie im SNF-Projekt „Merkmalsorientierte Diagnostik in adaptiven Lernsystemen (MODAL, Projekt-Nr. 1114-63536.00) Dissertationsschrift: Sag’ mir, was ich noch nicht weiß – Individualisierung des Wissenserwerbs über adaptive Auswahl der Lerninhalte mithilfe der Nonmetrischen Multidimensionalen Skalierung und Prokrustes-Transformation
1996-2002	Studium der Psychologie, Soziologie und Neurophysiologie, Universität Zürich Lizentiat: Juni 2002 Lizentiatsarbeit: Unterschiedliche Wissensqualitäten und ihre Auswirkungen auf Kognitive Strukturen beim Lernen
1989-1996	Gymnasium St. Antonius, Appenzell (AI), Matura Typus E
1983-1989	Primarschule in Schwende (AI)

Weitere Ausbildung

2006	Einführung in XML-basierte Frameworks (eLML), Universität Zürich, Informatikdienste
2006	12. Workshop für Angewandte Klassifikationsanalyse im Landhaus Rothenberge (Münster, Deutschland)
2006	E-Learning FORUM (SS06) an der Universität Zürich, E-Learning-Center (ELC)

2006	Blockkurs Strukturgleichungsmodelle (SEM), Universität Zürich, Psychologisches Institut
2005	didactica-Kurs „Kompetent prüfen – Neue Herausforderungen in gestuften Studiengängen“ an der Universität Zürich, Arbeitsstelle für Hochschuldidaktik (AfH)
2005	E-Learning FORUM (SS05) an der Universität Zürich, E-Learning-Center (ELC)
2003	SIDOS Summer School on Advanced Methods in the Social Sciences (Multidimensional Scaling) in Lugano

Beruflicher Werdegang

seit 2007	Leitung eLearning im Masterstudiengang Psychotraumatologie der Medizinischen Fakultät der Universität Zürich (http://www.mas.psychotraumatology.uzh.ch)
seit 2005	Projektmediator (Koordination) im Swiss Virtual Campus (SVC) Projekt „Psychopathology Taught Online“ (PTO) Projekt-Nr. 3-008, an der Universität Zürich, Psychologisches Institut, Angewandte Kognitionspsychologie (http://www.pto.uzh.ch)
2002-2004	Assistent an der Universität Zürich, Psychologisches Institut, Angewandte Kognitionspsychologie (inkl. Lehraufträge)
1999-2002	Wiederholte Semesterassistent, Universität Zürich, Psychologisches Institut, Allgemeine Psychologie (Kognition)
2001	Praktikum an der Abteilung Management und Leadership Development, Geschäftsbereich Human Resources, UBS AG, Zürich
2000	Praktikum am Institut für Marktforschung (ifm) in Freiburg i.Brsg. (D)
1998-1999	Interviewer für das Marktforschungsinstitut Demoscope in Adligenswil (LU)